Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

A Systematic Review of Explainability in Computational Intelligence for Optimization

Jose Almeida, Joao Soares, Fernando Lezama, Steffen Limmer, Tobias Rodemann, Zita Vale

2025

Preprint:

Copyright Elsevier 2025. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

A Systematic Review of Explainability in Computational Intelligence for Optimization

José Almeida[a], João Soares[a], Fernando Lezama[a], Steffen Limmer[b], Tobias Rodemann[b], Zita Vale[a]

[a]*GECAD - Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, LASI - Intelligent Systems Associate LAboratory, ISEP, Polytechnic of Porto, Porto, Portugal*
[b]*HRI-EU – Honda Research Institute Europe, Offenbach, Germany*

Abstract

This systematic review explores the need for explainability in computational intelligence methods for optimization, such as metaheuristic optimizers, including evolutionary algorithms and swarm intelligence. The work focuses on four aspects: 1) the contribution of Explainable AI (XAI) methods to interpreting metaheuristic performance; 2) the influence of problem features on search behavior and explainability; 3) the role of mathematical theory in providing transparent explanations; and 4) the potential of metaheuristics to enhance the explainability of AI models, such as machine learning (ML). XAI methods such as SHAP, LIME, and visualization techniques provide valuable insights into metaheuristic performance, while landscape analysis and quality diversity approaches reveal algorithm performance across different problem landscapes. The review also explores how metaheuristic algorithms can enhance the interpretability of ML models, turning black-box models into more transparent systems. The work moves on to proposing "Explainergy," a novel concept for integrating explainability into metaheuristic algorithms within the energy domain, enhancing the transparency and usability of optimization models.

This review is a foundation for future research combining explainability with evolutionary computation and metaheuristic optimization to address real-world challenges in diverse fields, including energy systems.

Keywords:
Artificial intelligence, Black box optimization, Computational intelligence, Explainability, Explainergy, Metaheuristics, Systematic review

Nomenclature

ABC Artificial bee colony

AI Artificial intelligence

ANN Artificial neural network

CA Cultural algorithm

CCEA Cooperative coevolutionary algorithm

CELA Context-aware evolutionary algorithm

CI Computational intelligence

CMA-ES Covariance matrix adaptation evolution strategy

DE Differential evolution

DL Deep learning

DM Decision-making/maker

DT Decision tree

DU Diversity and usage

EA Evolutionary algorithm

EC Evolutionary computation

ELA Exploratory landscape analysis

ES Evolution strategies

FCM Fuzzy cognitive maps

FEA Factored evolutionary algorithm

FJA Fast jaguar algorithm

GA Genetic algorithm

GD Great deluge

GP Genetic programming

GUI Graphical user interface

HC Hill climbing

IOA IMMa optimization algorithm

LLH Low-level heuristics

LLM Large language models

LMDS Landmark multidimensional scaling

LON Local optima network

MAP-Elites Multi-dimensional archive of phenotypic elites

MDS Multidimensional scaling

MFEA-DGD Multifactorial EA with diffusion gradient descent

ML Machine learning

MOEA/D Multi-objective evolutionary algorithm with decomposition

NSGA-II Non-dominated sorting GA

PopDP Population dynamics plot

PSO Particle swarm optimization

QD Quality diversity

REVAC Relevance estimation and value calibration

SB Structural bias

SI Swarm intelligence

SOMA Self-organizing migrating algorithm

TS Tabu search

V-MDF Visualizer for metaheuristics development framework

VRP Vehicle routing problem

WoS Web of Science

XAI Explainable AI

1. Introduction

The current environment in the optimization field is evolving with the increasing complexity of real-world problems. In this context, traditional mathematical optimization approaches may lose efficiency and efficacy, reaching their limit when dealing with extremely uncertain, nonlinear, high-dimensional, and constrained problems [1]. This lack of efficiency and efficacy can be more critical when optimization time is a prioritization due to the time horizon the problem is scaled to, for example, short-term problems where the solutions may be needed close to real-time by the decision-makers (DMs). As such, other optimization approaches have emerged to find suitable solutions to complex problems, raising them as effective problem-solvers. These approaches are computational intelligence (CI) search procedures that emerge from natural computing called metaheuristics or randomized optimizers [2]. These optimizers seek to emulate the nuanced problem-solving mechanisms found in the environment. These approaches encompass a wide range of techniques, from single-solution to population-based algorithms, where evolutionary algorithms (EAs) and swarm intelligence (SI) [3] are highlighted, all aimed at efficiently navigating complex problem spaces [4]. By mimicking the adaptive and exploratory behavior found in natural systems, metaheuristics offer versatile and robust solutions to a diverse array of challenging optimization problems since they include the need for less data and a priori assumptions to discover solutions for a given circumstance, flexibility in describing solutions, the ability to model any criteria in the form of fitness functions and providing a solution even if prematurely interrupted [5, 6].

1.1. Motivation

Metaheuristic optimization has increased and has gained significant traction in its implementation to optimize complex problems as they achieve a balance between "optimal" and "practical" solutions. Now, there are hundreds and even thousands of metaheuristic algorithms based on numerous

aspects of nature, and one concern is that experts are now more concerned with developing new metaheuristic algorithms than improving and tackling the challenges of what is already well-established. Plenty of metaheuristic algorithms are implemented without a concise implementation in real-world applications [7, 8].

As evidenced, metaheuristic optimization is a growing field, mainly concerning computer science and, recently, engineering, to tackle challenging optimization problems. This increasing popularity is theorized in [9] to be related to the high flexibility these algorithms provide through hybridization and customization since today's academic and commercial problems characteristics can change unexpectedly. Optimization methods need to quickly adapt, which are characteristics of a complex real-world problem. These problems are, in most cases, highly non-linear, with multi-modal landscapes in which generalized algorithms and structures are seeing their implementation for better DM [10]. Even though the use of metaheuristic optimization algorithms has been increasing to solve real-world problems and has shown their efficiency for large-scale optimization [6], many engineers, specifically in the authors' field, energy engineers have been reluctant to adopt these methods to deal with the increasing complexity of their sector. These concerns come from the major drawbacks associated with implementing these algorithms, which are the lack of guarantee of optimality and missing the explainability features about their performance and how they achieve the results they present since they are randomized optimization algorithms [11]. For each run performed, a new solution is given with a given noise, so they present concerns regarding its stability and robustness. The difficulty of understanding metaheuristic performance using solid mathematical foundations is one of the major disadvantages of implementing such methods. As such, a rigorous analysis of the algorithms' performance and search process should be taken into consideration, adopting intrinsic and/or post-host understandable explanations following the recent trends verified in the field of machine learning (ML) in artificial intelligence (AI), the so-called explainable AI (XAI) [12, 13], enabling an understanding of the algorithm itself or the problem features that influence algorithm behavior [14, 15]. The opposite can also be put into practice with metaheuristics explaining ML models, such as counterfactual explanations [16, 17] or genetic programming (GP) [18, 19]. These explanations would explain why the given solution was chosen, support decision-makers, and increase engineers' trust in such methods in industry and real-world applications beyond the academic world.

The intersection between XAI and metaheuristic optimization seems viable to provide the end-user with an easy-to-understand white-box model with practical and effective solutions for both complex problems and complex metaheuristic optimization and ML processes in which explainability is essential [20, 21]. With such a model, the term explainable metaheuristics can be introduced to bolster users' trust in these optimizers and methods.

1.2. Contributions

To tackle explainability in metaheuristic optimization, this work presents a systematic review of models and methods in the literature that attempt to explain such optimizers, with some contributions in utilizing metaheuristics to provide understandable and transparent ML methods. A thorough state-of-the-art analysis is made in this work, which aims to fill the gap in adopting such methods in real-world applications, mainly when it comes to metaheuristics. As such, the major contributions of this review work are summarized as follows:

- Provide a state-of-the-art analysis based on the PRISMA methodology [22] regarding the topics of explainability for metaheuristic optimization and its' sub-fields, elaborating multiple search queries for information gathering on the topics using different digital libraries;
- Gather explainable methods for metaheuristic optimization that are prominent in the literature based on theoretical backgrounds, XAI, and other different implementations that aim to explain and understand the performance of these optimizers;
- Summarize the use of metaheuristics to provide concrete explanations to complex black-box ML models, providing transparency in these mechanisms;
- Propose the "Explainergy" concept, which involves the development of an explainable, easy-to-understand metaheuristic framework for complex, large-scale optimization in the energy domain.

Fig. 1 shows the overall taxonomy of the approaches involving metaheuristic explainability considering the enumerated contributions of the proposed review work.

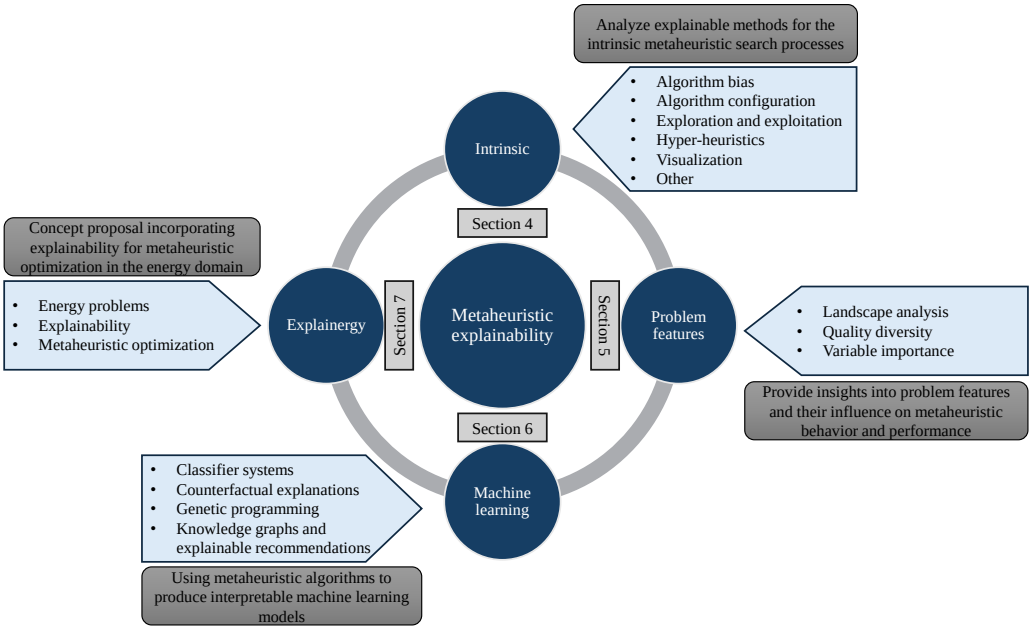

Figure 1: Overall taxonomy for metaheuristic explainability.

Even though this work presents some research regarding the utilization of metaheuristic algorithms to explain ML models, the research focuses on the opposite: ML models and theoretical mathematical explanations are used to understand metaheuristics' performance and search process.

1.3. Paper structure

The manuscript is structured as follows: Section 2 introduces explainability and its terms with the target audience for metaheuristic explainability approaches. Based on the PRISMA tool, the review methodology regarding the information obtained from different digital libraries for given search strings is presented in Section 3. Section 4 presents the main subjects taken from the literature concerning intrinsic metaheuristic explainability, and Section 5 gives approaches that provide problem insights influencing metaheuristic optimization. Section 6 shows how metaheuristics can offer explainability to ML models. A discussion regarding the future of this topic and the steps needed to increase further interest regarding the application of these methods in real-world applications are presented in Section 7. Also, this section introduces a concept called "Explainenergy," involving explainable EC methods in the energy domain. Finally, the main conclusions are drawn in Section 8.

2. Explainability

Regarding the use of explainability in AI, named XAI, two major questions are evoked: What is it, and who is it for? In the field of XAI, multiple terms emerge to provide insights into AI models, namely ML forecasting analysis together with *explainability*, such as *understandability*, *comprehensibility*, *interpretability*, and *transparency*. The survey work in [13] clearly defines these terms and delineates *understandability* as the most crucial term in XAI with *transparency* and *interpretability* being tied to this concept. Still, in this work, *explainability* is used as a general term since it depends on the proposed method and its *target audience*. The *target audience* can vary across users based on their experience, background, and acquaintance level with the AI system.

Metaheuristics are mostly AI-inspired approaches based on nature-inspired evolution with large applicability, mostly for optimization problems that explore areas of the search space through solution-based approaches and variation operators that introduce diversity in this exploration. In further sections, metaheuristic optimization is used and described for producing interpretable or symbolic models for classifier systems (Section 6.4) and genetic programming (Section 6.2), where they can produce explanations for these AI systems. In contrast, integrating XAI into metaheuristic optimization can help explain the decision-making process of these algorithms, which facilitates debugging and refining them, where non-expert or non-technical DMs, for instance, engineers or medics, where solution understanding and trust are essential.

As such, for an AI system and making its connection to the world of metaheuristics, which are AI-inspired randomized optimization methods, three different entities that explainability can be for are proposed, as Figure 2 demonstrates. The first entity refers to the metaheuristic developer, who, in most cases, is a computer scientist or software expert with deep insights into the metaheuristic search processes. In contrast, explainability refers to methods that help design and understand these metaheuristics for performance improvements. The other entities are regulatory agencies, which are government agencies (lawmakers) that need to validate the model's compliance with the imposed regulations through system stability, reliability, fairness, and bias. The third and final entity that metaheuristic explainability can be for are naive users with a very limited understanding of these algorithms and no inclination to dive deeper into the intrinsic metaheuristic search mecha-

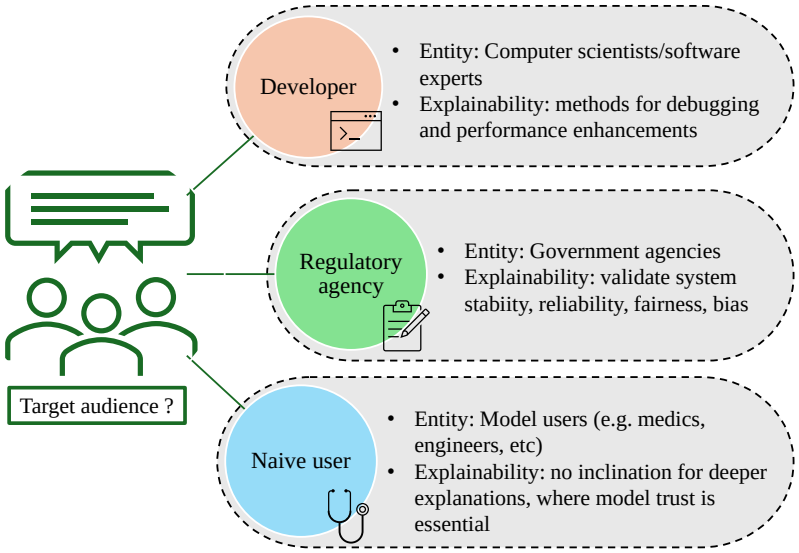

Figure 2: The target audience for explainability methods in metaheuristic optimization models.

nisms. Here, psychological factors become more important, where the feeling of trust in the given model is essential.

In this context, when focusing on metaheuristic algorithms, the metaheuristic explainability approaches as *intrinsic explainability* and *problem feature explainability* are grouped compared to [21], which proposes interpretability by design, explaining problem landscapes, explaining algorithm solutions and behavior. As such, the proposed *intrinsic explainability* aggregates both explaining algorithm solution and behavior and interpretability by design, where explaining problem landscapes is directly associated with this manuscript's proposed *problem feature explainability*. Concerning intrinsic metaheuristic explainability approaches, they are intended to provide explanations to the metaheuristic search procedures through bias explanations, exploration, and exploitation analysis, which are the processes by which metaheuristics get to a solution and visualization methods, for example. Regarding problem feature explainability, it is important to understand which problem features influence metaheuristic behavior, which can help to improve metaheuristic design and even problem formulation.

The *target audience* is also considered when referencing the different works for each category, where mostly the developer is where those explana-

tions are for, since most of the works focus on more detailed insights that allow a deeper understanding of the algorithms for a better debugging and to improve metaheuristic performance. Still, it should be aimed to produce an understandable algorithmic behavior to provide trust and transparency in these methods, mostly to non-experts who need to understand and interpret the given solutions since most DMs are unfamiliar with these algorithms. For example, it is very common in the energy field, where engineers and DMs are not conformable and do not understand these CI optimization methods. The introduction of an explainable energy framework ("Explainergy" in Section 7) could provide significant insights into these methods for a vast number of problems, which in many cases are large-scale, such as the energy resource management, optimal expansion planning of large energy networks, among others.

3. Review methodology

The current study was conducted largely following the PRISMA 2020 systematic literature review method [22]. This methodology provides a standardized framework that ensures clarity, consistency, and completeness in reporting systematic review processes and findings, enhancing this work's utility. By following PRISMA 2020, this study maintains methodological rigor and transparency, facilitating critical appraisal and enhancing the reliability of its findings for evidence-based DM. In this section, this framework is intended to be followed to its' maximum, where the research topics related to metaheuristic optimization are enumerated. The considered digital libraries and elaborated research queries are also detailed, along with the information retrieved from the queries.

3.1. Research topics

The use of metaheuristics is gaining notoriety in the world of optimization. These are adaptable (can be used to optimize any problem), simple to implement, and relatively efficient in solving complex, non-linear, high-dimensional, and highly constrained problems. These algorithms, also called randomized optimization algorithms, rely on random search techniques to find a global optimum solution to the problem(s). They present exploration and exploitation processes through diversification and intensification of the search space [3]. Diversification is used in the initial phase so the algorithm

can explore the search space to find diverse solutions and intensify it in the later stages, increasing exploitation when good solutions are found.

To achieve these diverse mechanisms to explore the search space, metaheuristics base their behavior on everyday nature-inspired mechanisms regarding human and animal behaviors and physical processes. These algorithms can then be divided into multiple sub-fields such as EAs, SI-based algorithms, human-based algorithms, and physics-based algorithms, as Fig. 3 shows [3, 4]. Inside these categories are numerous optimization methods, which nowadays go to thousands or dozens of thousands, which would be impossible to enumerate in the figure. Instead, in the presented representation, some of the most popular methods are represented inside the corresponding category. Some experts in the field do not agree with the proposed metaphor-based categorization in the figure, where some of these metaheuristics are not well accepted in the metaheuristic community since most are not reinventing the wheel and not proposing novel approaches compared to earlier and well-established approaches such as Tabu search (TS) and Simulated Annealing (SA), even though the authors claim they are [7]. Still, this categorization is presented since these algorithms are extremely popular, with large applications including the Grey Wolf optimizer [23] with over 16,000 citations and Cuckoo search [24] with over 8,000 citations.

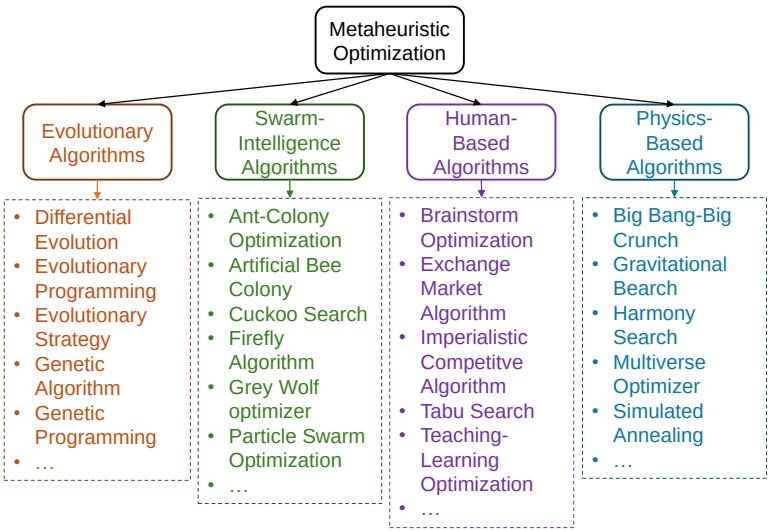

Figure 3: Classification of metaheuristic optimization methods, adapted from [3, 4].

3.2. Digital libraries and search queries

Different digital libraries were utilized to obtain relevant publications regarding the research topics previously introduced. Scopus, Web of Science (WoS), and the IEEE Xplore digital libraries were selected to retrieve the most relevant information. IEEE Xplore only includes publications from their publisher and corresponding area, which can give more detailed information regarding computing and engineering, but the information obtained is less diversified. On the contrary, Scopus and WoS are multidisciplinary digital libraries, containing not only scientific publications in computing and engineering, as the other two digital libraries, but from all scientific disciplines, enabling a more thoughtful review of the pertinent research topics posed. Regarding the terminology used to formulate the search queries, the different metaheuristic optimization categories were considered, as well as some of the most popular algorithms inside those categories together with the explainability and its variations, as section 3.1 showed. As such, the terms used are as follows:

- MT = explainable OR explainability OR explanatory OR explanation
- T1 = metaheuristic* OR meta-heuristic* OR meta heuristic*
- T2 = swarm intelligence OR swarm-intelligence OR swarm-based algorithm* OR swarm based algorithm*
- T3 = evolutionary algorithm* OR evolutionary-algorithm*
- T4 = human-based algorithm* OR human based algorithm*
- T5 = physics-based algorithm* OR physics based algorithm*
- T6 = particle swarm optimization OR particle-swarm-optimization OR ant colony optimization OR ant-colony-optimization
- T7 = genetic algorithm* OR genetic-algorithm* OR differential evolution OR differential-evolution OR genetic programming OR genetic-programming
- T8 = tabu search OR tabu-search OR cultural algorithm* OR cultural-algorithm*

- T9 = simulated annealing OR simulated-annealing OR harmony search OR harmony-search

The main term includes "explainable," "explainability," "explanatory," and "explanation" since these terms represent great value for the work conducted regarding explainability in metaheuristic algorithms. The other terms encompass the metaheuristic categories and algorithms, as previously explained. Different operators were used regarding query construction. The '*' was used in "Metaheuristic*," "Meta-heuristic*," "Evolutionary algorithm*," "Genetic algorithm*," and others to cover their plural forms. Using operands such as AND and OR allowed the concatenation of different terms to put together different queries.

The first query intends to encompass broader search areas by including only the term associated with metaheuristics. The following four queries include this term and the different enumerated categories. The final four categories intend to be more specific by incorporating specific algorithms for each category. The query pipeline can be observed in Fig. 4.

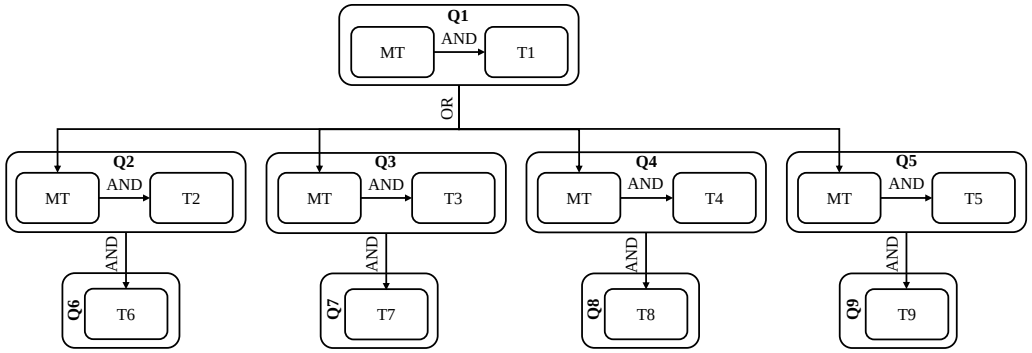

Figure 4: Pipeline representing query formulation.

Web-based search tools provided by the selected digital libraries empower researchers with the formidable capability to recreate or duplicate the original study, effectively replicating the insight or knowledge encapsulated in the research.

3.3. Information gathering and Synthesis

Table 1 lists the number of publications found in each digital library for each separate search query and the total number of publications in

Table 1: Number of publications retrieved in each digital library for each search query.

Digital library	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Q7**	**Q8**	**Q9**	**Selection**
IEEE Xplore	52	71	87	50	50	28	28	1	5	117
Scopus	205	106	415	0	0	29	139	0	0	681
WoS	105	132	285	105	105	110	164	105	105	311
$\sum Q^1$	362	309	787	155	155	167	331	106	110	1,109
$\sum Q^2$	258	268	601	142	143	159	283	106	108	797
Duplicated	104	41	186	13	12	8	48	0	2	312

[1]: with duplicates; [2]: without duplicates

each database, including the number of duplicated publications and non-duplicated ones. Note that the number of duplicated publications comes from each database, where all queries were concatenated with the operator OR to eliminate duplicated publications. The bibliometrix[1] package was used to identify the number of duplicate publications in each database query. By excluding all duplicated publications between queries and between digital libraries, the final number of publications to analyze was 797.

The publications obtained after the duplication-removing process were then analyzed, and selection criteria regarding three different research questions were taken into account to filter further the number of publications in the scope of this work. The research questions considered are as such:

- RQ 1: Do XAI methods contribute to explaining metaheuristic performance behavior?
- RQ 2: Do problem features influence metaheuristic diverse search procedures, and can their analysis have greater explainability?
- RQ 3: Can theoretical mathematical backgrounds provide understandable and transparent explanations for metaheuristic optimization?
- RQ 4: Can metaheuristics be utilized to introduce explanations in AI methods, namely ML models?

In order to respond to these questions, the total number of publications retrieved from each database was analyzed according to the process in Fig. 5

[1] `https://www.bibliometrix.org/home/`

following the PRISMA methodology. Initially, the title, abstract, and keywords were scanned to see if they matched the topics approached in the manuscript and if they were irrelevant to the research questions posed. By filtering, many publications were excluded from each database, resulting in 485 publications removed from further analysis, with 312 remaining. Upon thoroughly reviewing the remaining publications through a full-text analysis, 187 studies that did not meet the requisite inclusion criteria were excluded. Other 26 publications considered relevant for the topics addressed in this review were manually added for analysis upon examining some of the bibliographies of the 125 primary studies, resulting in 151 distinct papers constituting the comprehensive collection of primary studies.

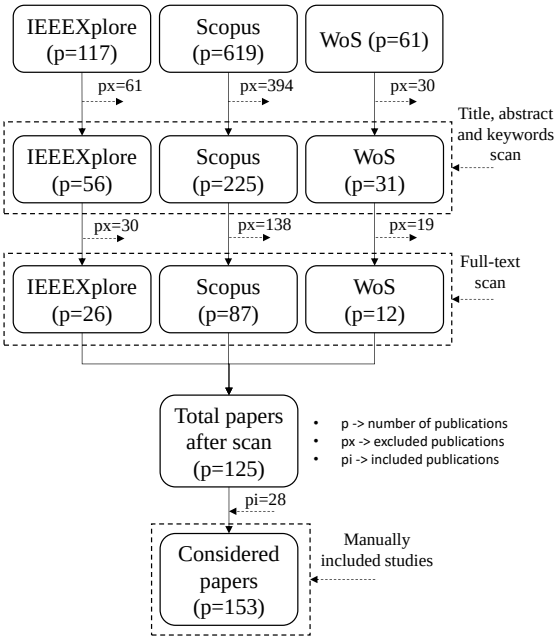

Figure 5: Primary studies selection process regarding RQ 1, RQ 2, RQ 3, and RQ 4.

Fig. 6 shows the number of publications per year from the selected primary studies. The selected studies date back from 1997, with a very low number of publications until 2017, ranging from 1 to 3 publications a year, with years without publications. From 2017 onward, an exponential increase has been verified, primarily due to the scorching topic of AI and, more precisely, XAI, which has been increasing in notoriety in recent years. A drop from 2022 to 2023 and 2024 is expected since this review was conducted

at the beginning of 2024, and more publications may appear regarding this topic until the end of the year, with some that still may appear from 2023. Still, from the figure, the topic of explainability concerning metaheuristic optimization has gained traction in the past few years.

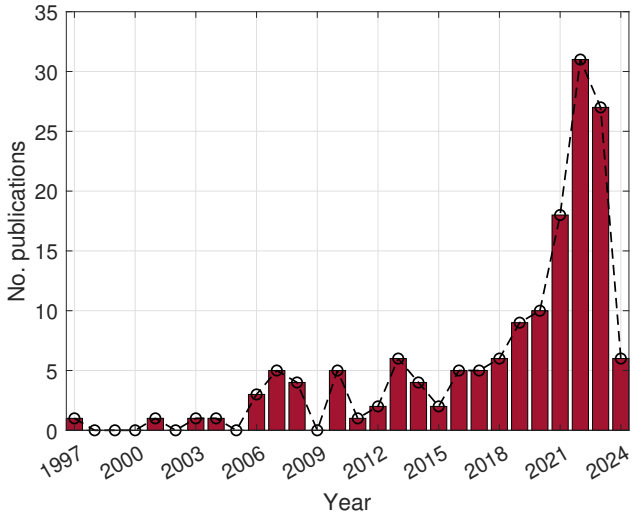

Figure 6: Number of scientific publications per year from the primary studies retrieved.

Note: The presented search methodology was conducted in February 2024. A new search would change the entire methodology process.

4. Understanding metaheuristic algorithms

Concerning the information gathered from selected primary studies, different methods that attempt to provide intrinsic explanations to metaheuristic algorithms were gathered, and Table 2 shows these different methods together with the used metaheuristic category and the total number of publications. From the table, such methods intend to provide explanations for internal metaheuristic search processes and their hyperparameter behavior analysis, complemented with fundamental theoretical explanations.

This section further explains the summarized works in the table, provides detailed information, and presents the main takeaways from each manuscript.

Table 2: Taxonomy of the intrinsic explainability methods for metaheuristic optimization.

Method	**Category**	**No. Pubs**	**References**
Algorithm bias	EAs	8	[25, 26, 27, 28, 29, 30, 31, 32]
Algorithm configuration and design	EAs	9	[33, 34, 35, 36, 37, 38, 39, 40, 41, 32]
	Human-based	4	[42, 43, 44, 45]
	SI	3	[46, 47, 39]
Exploration/exploitation analysis	EAs	5	[48, 49, 50, 51, 52, 41]
	SI	3	[50, 53, 54]
Hyper-heuristics	EAs	3	[55, 56, 57]
	Physics-based	1	[58]
	Quality-diversity	1	[59]
	SI	1	[57]
Visualization	EAs	7	[60, 61, 62, 63, 64, 65, 66]
	Human-based	1	[42]
	SI	2	[60, 65]

EAs: evolutionary algorithms; SI: Swarm intelligence.

4.1. Algorithm bias

Metaheuristic optimization performance evaluation under different conditions is essential in determining whether an algorithm is suited for a single problem or a multitude of problems. As such, to understand algorithm behavior and other additional information in these situations, algorithm bias or, in some particular cases, structural bias (SB) analysis can be introduced. Metaheuristic bias towards a certain area in the search space becomes interesting to analyze, so explanations can be given to understand such behaviors. An early work in [25] addresses this concept by exploring the skewness of mutation operators as a factor of bias. Through visualization techniques such as genotype-phenotype mappings, the authors expose the biases and counterintuitive behavior of two simple mutation operators (normal and flat-balanced) to improve the understanding of these genetic encodings and create helpful explanations that will help a better understanding of the more intricate genetic encodings. Another work proposes biasing cooperative coevolutionary algorithms (CCEAs) towards the best possible cooperation [26]. A degree of bias results from the influence of a weighted sum of two rewards

in the objective function, where the bias, problem features, and algorithm features are investigated. It was found that some problem features significantly affect the prior static biasing strategy, where a stochastic alternative that assigns two different performance evaluators per individual mitigates this issue. The Edge-Window-Decoder [27] is a method that intends to reduce problem complexity, mainly on combinatorial optimization problems with tree-based structures, and enhances metaheuristic performance. This approach has high locality and heritability and an intermediate level of representational bias toward the minimum spanning tree problem. Even though the basic Edge-Window-Decoder approach is unbiased, its good performance on such problems may be attributed to the inherent bias associated with the genetic operators of this method. The genotype-phenotype mapping inherently generates biases that affect the thoroughness and efficiency of exploring certain areas within the search space, which is explored in [28] for GP. The research examines how genotype-phenotype mapping in GP brings intrinsic biases into the search process, influenced by neutrality, robustness, and network architecture. These biases influence the evolutionary processes and impact the efficacy and results of GP-based searches.

More recent works consider this term of SB to describe the inherent tendencies of metaheuristic algorithms. Metaheuristic algorithms suffer from SB, which degrades the algorithms' performance in the objective space. SB may determine which situations are less likely to result in SB and pinpoint areas for improvement by recognizing its kind, strength, and existence in a heuristic optimization method [31]. There are some common types of SB, such as bias towards the center of the search space, bias towards the bounds of the search space, bias towards certain parts of the search space forming clusters, bias towards avoiding certain parts of the search space, creating gaps, and strong discretization [31].

Tools for detecting SB are few, but in [29, 67, 30], such tools are presented, where an analysis is made regarding the effect of SB in metaheuristic behavior. The work in [29] goes further and analyzes the anisotropy and its relation to SB, where an algorithm that exhibits anisotropy has SB but, in some cases, shows only SB in some of the dimensions of the problem. As such, the authors discover that anisotropy is extremely uncommon, and in a few instances, it does occur. In fact, anisotropy presents no negative impact on the detection of SB. The authors also propose that understanding the different kinds of SB would allow a more robust plethora of testing procedures to detect any sign of SB covered in [31]. This work proposes a

toolbox for SB, integrating XAI under the name Deep-BIAS. This toolbox uses a trained deep-learning model that instantly determines the kind and strength of SB based on the raw performance distributions. Different metaheuristics are tested, such as variations of DE, compact solution algorithms, and single-solution algorithms. The most prominent type of SB detected was the bias towards a center of the search space, and the integration of XAI enabled this understanding. Furthermore, in [30], the behavior of the CMA-ES metaheuristic is analyzed for many configurations for the BBOB benchmark functions. More than 90% of the algorithm configurations used presented some SB, with more than 50% being biased towards the search space's center. Recently, a new framework for explainable benchmarking in metaheuristics has been proposed in [32], incorporating an SB analysis. This analysis considers the single-run best and average best configurations of a Modular CMA-ES and DE with the type of bias that the toolbox obtains for the COCO benchmark functions with different dimensions. The conclusion from this work regarding the SB is that the tested Modular CMA-ES for almost all configurations selected and for low dimensional problems presents a bias towards the center of the search space, where the Modular DE moves towards the bounds of the search space but exhibits other types of SB more frequently being not as uniform as Modular DE.

Detecting any type of bias is key to understanding metaheuristic behavior, and the tools are necessary to explain the type of bias the metaheuristic presents. Still, even though some works intend to explain why this bias is being observed, many improvements can be made regarding the literature and the implementation of such tools and mechanisms in real-world problems.

4.2. Algorithm configuration and design

Due to their stochastic nature, the proper design of a metaheuristic algorithm is important so maximal robustness for a wide variety of problems can be obtained with acceptable solutions. Understanding the behavior of the parameters associated with a specific metaheuristic is important in comprehending its behavior and explaining metaheuristic performance. Even in some works, the authors change the design of the algorithm to make its behavior more explainable by itself.

The works in [38, 40] exemplify these modifications perfectly, considering the nature of the Fast Jaguar Algorithm (FJA), which the authors claim to be explainable in their design. This EA is claimed to be transparent and interpretable, presenting explanatory characteristics as in [38]. The algorithm uses

historical information to determine its course in the search space and find the global optima effectively. An improvement of this algorithm by incorporating a precision hunting mechanism is proposed in [40]. This mechanism allows the algorithm to accurately seek prey and adaptively identify patterns in the present environment while moving, not using random procedures when going through the search space, making the path of the algorithm interpretable since it gives the same outcomes when using the same inputs. The work in [46] also improves the standard Artifical Bee Colony (ABC) metaheuristic to make it more explainable for optimizing multiple benchmark functions. The authors integrate an algorithm called β-hill climbing (βHC); therefore, the algorithm is called ABC–βHC, making the local search process of the ABC substantially better and improving the overall exploitation. This local search method is an information-based procedure that eliminates the randomness associated with ABC in the selection process. This process makes the algorithm more explainable and interpretable to the developers, with its statistical significance compared to other metaheuristics.

Similarly, [43, 44] design a trustable framework for Cultural Algorithms (CAs). In both works, the authors use the Common Value Auction (CVA) configuration to support fairness, explainability, responsibility, and sustainability of the proposed framework, and a comparison is made in both works with the Wisdom of Crowds configuration. The first configuration presents higher sustainability when compared to the other configuration, regarding resilience and robustness of the algorithm, since in this configuration, the information for each of the agents in the metaheuristic is higher. Indeed, with this configuration, the authors could satisfy the property for an XAI algorithm, which is a major step towards designing algorithms that are easily understood and interpretable by users.

Solution feasibility in metaheuristic optimization is an important step to guarantee the efficiency and efficacy of these algorithms. A pragmatic approach involves creating a possible solution and, if required, converting it into a feasible one; ideally, this process would provide a local-optimal solution. Set theory aims to correlate an element of the solution set with an element of the feasible solution set, allowing for analysis and optimization from a novel perspective [39]. Mathematical explanations of the solution transformation process are given in [39] based on mapping relations and the theory of preimages to improve and provide satisfactory properties the operators need. The work shows that the number of preimages and their size directly relate to the metaheuristic efficiency, progressing the literature in metaheuristic design to

guarantee solution quality and computational efficiency.

Other works opt to evaluate the metaheuristic parameters, seeing their influence on the algorithms' performance, and try to provide explanations for such behaviors based on the tuning of the parameters. Early works such as [42, 33, 34, 35, 36] exemplify this. For instance, in [42], a Visualizer for Metaheuristics Development Framework (V-MDF) tunes a TS algorithm for combinatorial optimization. The framework incorporates user interaction, allowing users to determine effective search strategy adjustments by observing operation results stored in a rule base. In [33], the behavior of the crossover probability parameter in DE is studied alongside move magnitude and scale factor influences. It reveals that low crossover values lead to frequent minor improvements across individuals, while high values drive occasional but significant enhancements, often reducing population size and benefits over time. These findings emphasize the balance required in parameter selection. The third work [34] proposes a Relevance Estimation and VAlue Calibration (REVAC) method for many problems, enabling the tested EA to be more robust and more general through the fine-tuning of the parameters. The REVAC mechanism falls into an iterative generate-evaluate heuristic search-based method [15]. It concludes that the more generalist parameters perform far worse than those obtained for a specific problem. Choosing parameters that perform well for multiple problem instances can be a challenging task that still needs attention in the literature. Looking at [35], a self-adaptive automatic DE parameter tuning is proposed, with no human input, introducing a mutation operator that addresses the issue of stagnation in local minima. The IMMa Optimization Algorithm (IOA) showcases consistent optimal performance across various cases, whether requiring precision, following a trajectory, or incorporating time constraints. In [36], how asymmetric neighborhood assignment affects the performance of a multi-objective EA based on decomposition (MOEA/D) is explored. The statistical analysis demonstrates the influence of neighborhood size on Pareto front localization. The study concludes that odd neighborhood values mitigate asymmetric convergence behavior, improving algorithm reliability.

More recently, [37] evaluates the performance of a Factored Evolutionary Algorithm (FEA) for many parameter values applied to deceptive unitation functions and different instances of the Royal Road problem. The authors experimented with the factor size and overlap parameters. They concluded that for the unitation functions, the Compete step in FEA (subpopulations with the state of best fitness) encounters difficulties in generating optimal

solutions. Regarding the Royal Road functions, the most effective factor architecture directly links factors to schemata (elements directly from {0, 1,#} that denote a pattern in a bit string).

In other ventures of algorithm configuration, [45] also uses an iterative generate-evaluate heuristic search-based method [15], the iterated F-race method for multiple metaheuristic configurations for different problem instance sizes. Instead of tuning parameters, the authors configure linear and piecewise-linear models representing parameter values. These configurations outperform those from the original design method, even with limited setup resources. Using visualizations and simulations, [47] examines the Self-Organizing Migrating Algorithm (SOMA), focusing on the perturbation operator. Lower perturbation slows diversity loss, extends convergence time, and limits individual migration, while higher values reduce search space exploration within migration bounds. The described framework in [32] explores CMA-ES and Modular DE configurations, where an approximation SHAP method, which uses less computational effort than the exact SHAP value, is employed to extract the contributions of each hyperparameter to the performance of the tested metaheuristics. The framework also uses an anytime performance measure, the normalized Area Over the Convergence Curve, to compare and analyze the performance of each algorithm configuration, demonstrating the importance of hyperparameter insights and offering improved explainability and efficiency in algorithm configuration.

A leader-follower approach in [41] works as a "Team" to optimize solutions. Every algorithm helps find solutions by expanding the search field or honing down on potential candidates. The authors provide an explainability framework where the end-user can understand the metaheuristic and its solutions. The important components are the hyperparameter impact, the average number of feasible/infeasible solutions, and solution diversity. Hyperparameters are adjusted using the Taguchi technique, a statistical tool for robust design. The performance and explainability of the suggested method are demonstrated by analyzing the exploration and exploitation behaviors of these parameters among follower metaheuristics. This improves comprehension and optimization effectiveness, showing its explainability.

In summary, regarding metaheuristic configuration and design, it is essential to understand and explore the parameters' influence and the metaheuristic itself to understand the algorithms' performance. This allows the developers to explain and modify the metaheuristic search process so its behavior can be explanatory.

4.3. Exploration and exploitation analysis

In metaheuristic optimization algorithms, exploring and exploiting a search space are crucial to producing good results. Exploration guarantees a more diverse search for solutions, allowing the algorithm to avoid being trapped in a local optima. Even so, when good regions of the search space are found, exploitation allows the metaheuristic to further dig into that region so better solutions can be found. So, understanding and providing explanations for these mechanisms is fundamental to comprehending and trusting metaheuristic behavior and performance.

Early works in [48, 49] present proximity and spread response maps, which show the behavior of multi-objective EAs for problem optimization with multiple conflicting objectives. An analysis given by these tools gives insight into how well the used MOEAs can balance exploration due to the spread of solutions and exploitation, given the proximity to the Pareto front. The authors concluded that the optimal point for achieving high-quality approximation (exploitation) tends to decrease as the number of conflicting objectives increases and that diversity promotion mechanisms (exploration) are highly influential and should be considered requiring some attention since it can pose a difficult challenge due to the poor approximation that comes with high diversity or spread. As such, finding the sweet spot between exploration and exploitation must be achieved, and it is of great interest when designing and evaluating such algorithms.

The work in [53] extensively analyzes the exploration and exploitation mechanisms for a diverse portfolio of swarm-based metaheuristics in a real-world scenario, where [54] explores the balance between exploration and exploitation in an artificial bee colony (ABC), explaining its behavior for high-dimensional problems. Both works intend to provide explanations for the exploration and exploitation behavior of swarm metaheuristics, wherein the first, an analysis of the behavior of each individual of the swarm as well as the swarm as a whole is made, and a visualization of the average exploration and exploitation for each algorithm is shown for each test instance so explanations of this behavior can be formulated. The second work concludes that one operator of the ABC algorithm called scout bee poorly performs for larger dimensional problems, which explains the poor exploration of this algorithm in such instances. In [50], the authors only explore the effect of the exploration mechanism in DE and PSO. This understanding in the analyzed works allows domain experts to properly configure such metaheuristics so

exploration and exploitation can be done effectively and good solutions can be found.

More recently, the works in [51, 52] attempt to explain metaheuristics' exploration and exploitation process through a recommendation model for the first work and transition information for the second work. In detail, [51] presents an influence analysis mechanism with its own variation operators separated from the algorithm influenced in this process. Multiple indicators that try to explain the exploration and exploitation processes are put in process as well, such as constraints completion: the average between constraint score and constraint maximum; exploration and exploitation power: given by the number of exploration, exploitation vectors, and digits that recently are passed to 1 and others that already passed to 1; temporality: average of the difference between maximum and minimum temporal gaps. The authors found that the influence process positively impacts the algorithm and does not induce randomness in the search process. Similarly, [52] intends to provide intel on how a metaheuristic utilizes the search space through a representation of presence variation vectors that make the transitions between exploration and exploitation clear to understand. This transition is similar to the previous work, where digit transition information is considered regarding the shifts between exploration and exploitation. Metaheuristics and test instances can be compared using the proposed transition indicators.

Diversity is a good measure of understanding and knowing regarding metaheuristic explainability, which is directly influenced by the algorithm's exploration capabilities. The work in [41] dictates that this behavior must be known in their explainability framework. As such, the authors analyze how different algorithms in the "Team" contribute to exploration and exploitation tasks, showing the behavior in this regard for the tuned hyperparameters. Diversity is essential for EAs to prevent early convergence and stagnation, which the authors evaluate through solution feasibility and several function evaluations of a solution, giving insights into the exploration dynamics of the used metaheuristics for complex problem optimization.

4.4. Hyper-heuristics

To tackle computationally challenging problems, a high-level automated search approach known as a "hyper-heuristic" searches a search space containing low-level heuristics, such as move or neighborhood operators, metaheuristics, or heuristic components. Hyper-heuristics may be divided into two

categories: those used to produce heuristics and those used to choose heuristics [68]. As such, with the intel from the "hyper-heuristics," metaheuristics and other heuristics can be chosen or produced for a given problem that better suits them, providing an insight into metaheuristic performance. In [68], a classification of hyper-heuristics is made, where they separate these algorithms based on feedback, low-level heuristics (LLH), solutions, objectives, move acceptance, and parameter setting. The authors characterize online and offline learning as mixed and no learning hyper-heuristics in the feedback classification. In the LLH, there is a type: construction or perturbation; a set: whole, reduced, and increase; and a grouping of LLH: without distinction, predefined sequence, and stage-based. The solutions are single-point, multi-point, and mixed, corresponding to the objectives encompassing single and multi-objective. The move acceptance corresponds to the DM process in the heuristic selection, which can be stochastic or non-stochastic, depending on whether a probabilistic framework is in place. The final classification depends on the parameter setting, which can be static, dynamic, adaptive, or self-adaptive.

Table 3 shows the classification of the hyper-heuristics used in the works obtained in the query process. The work in [55] proposes a hyper-heuristic for multi-objective optimization that uses a mixture of experiments through four rank indicators, allowing the algorithm to adapt regardless of the scenario. These rank indicators explain the influence of such metrics in the search and DM process, helping practitioners understand which heuristic combinations are most effective under the different scenarios. In [58], a hyper-heuristic approach is used for a multi-objective vehicle routing problem (VRP), where an objective related to the balance of distances between routes is added to the common total distance minimization. By considering this balance, DMs can understand how the algorithm prioritizes and balances conflicting objectives. The LLH model implemented and compares the different heuristics, adding an interpretability aspect by revealing which strategies perform better and why. In both works, using hyper-heuristics to select or generate LLH allows for a clear understanding of which LLHs are being used and why they are selected at different stages of the optimization process. The hyper-heuristic DM process can then be traced and analyzed, making it easier to explain algorithm behavior.

A merge of quality diversity (QD) optimization for sequence-based selection of hyper-heuristics is made in [59]. In the context of hyper-heuristics, this work incorporates a QD metaheuristic to provide information about

Table 3: Hyper-heuristic categorization.

Ref.	Year	Feedback	LLH set	Solution	Obj.	Acceptance	Parameter setting
[55]	2013	offline	reduced	MP	MO	stochastic	static
[58]	2019	mixed	reduced	MP	MO	NS	adaptive
[59]	2021	online	whole	MP	SO	NS	static
[56]	2024	mixed	whole	MP	SO	stochastic	adaptive
[57]	2024	mixed	reduced	MP	SO	NS	adaptive

MP: multi-point; MO: multi-objective; SO: single-objective; NS: non-stochastic.

hyper-heuristic performance when different instances are in place. The authors demonstrated that not only did the knowledge gained from the QD algorithm allow for a better understanding of the effect of multiple parameters in hyper-heuristic selection, but a comparison with LLH, randomly selected sequences, and a synthetic oracle was not able to outperform their proposed method for multiple sets of the balanced partition combinatorial optimization problem.

The use of large language models (LLMs) for heuristic generation and design has also emerged in [56, 57]. In particular, [56] proposes that heuristic concepts are articulated in plain language ("thoughts") and subsequently converted into executable code by LLMs. The framework develops "thoughts" and codes through an evolutionary search, creating high-performance heuristics. The work in [57] uses so-called reflections to guide the search. It utilizes LLMs to generate and adapt heuristics, and by incorporating reflections, they generate and modify heuristics dynamically based on past performance. These could be interpreted as explanations giving insights into the workings of heuristic generation and selection. Due to the nature of LLMs and their ability to generate outputs understandable by humans, both works have an intrinsic explanatory value by combining LLMs with EC, expressing heuristics in natural language. This integration increases transparency for the mechanisms behind the evolutionary processes, particularly in how heuristics evolve, adapt, and improve.

This topic presents a good research venue concerning explanations for metaheuristic tuning, selection, and building an algorithm portfolio by providing simpler, more generalist, and understandable hyper-heuristics suitable for diverse optimization functions across multiple instances, demonstrating their robustness and explainable behavior.

4.5. Visualization

In the context of XAI, visualization methods are a pillar that enables the interpretability of black-box optimization methods such as metaheuristics. They show how the algorithms attain a certain solution by presenting the metaheuristic optimization process, demonstrating and justifying the performance attained by these algorithms in a certain environment [69].

The work in [42] previously referenced uses a visual diagnosis for TS tuning. Regarding the visualization method, the work uses a visual diagnosis, proposing incidents as the combination of basic events such as local optima and an uphill/downhill movement, for example, to visually give the state the algorithm is in its search trajectory. It uses cause-action-outcome rules to respond to positive and negative incidents. It uses a V-MDF tool that implements distance radar plots and stores the obtained cause-action-outcome rules, as described previously, effectively helping the user discover and rectify negative incidents through proper remedial actions. In [60], performance profiles are used to visualize and compare the performance of different metaheuristics in the 2006 CEC benchmark functions. The authors present different performance measures extracted from these performance profiles (success rate, feasible rate, among others), which are compared in a parallel coordinates graph for each tested metaheuristic, enabling a better understanding of the ranking obtained. With the representation of these performance metrics, the interpretation and communication of intricate data pertaining to the execution of metaheuristic algorithms can be facilitated.

More recently, in [63], multidimensional scaling (MDS) and landmark MDS (LMDS) are utilized for visualization of the evolutionary process in the search space regarding the evolution of the populations for multiple many-objective EAs. The LMDS addresses the computational difficulties associated with visualizing many objective and complex issues in MDS. Exploration and exploitation visualizations are shown for a reduced objective space regarding the MDS and LMDS mechanisms, where a plethora of information can be obtained concerning EA performance, the timing and locations of population exploration/exploitation, and the specific regions of the search space that the population is perturbing. Similarly, in [62], a Population Dynamics Plot (PopDP) is proposed to enhance metaheuristic understanding by presenting a user-friendly visualization application that can represent the entire evolutionary search process of the NSGA-II metaheuristic for a multi-objective knapsack problem taking the previous work as a basis. It shows the population of solutions, the genealogy of parent-offspring, solution perturbation

operators, and the journey of the search process by maintaining a comprehensive log throughout the optimization process, detailing the generation process and the specific approach used to produce each solution and recording to ascertain the parent-offspring associations of solutions. It is stated that PopDP directly projects the solution operators, fitness, and lineage; DM can be guaranteed with this information. Still, these works assume that recipients have an understanding of how metaheuristic algorithms work for the DM, so a survey for non-experts and users who need to comprehend how metaheuristics work should be put into practice to understand the usability of these models.

A work that takes into account this last statement and considers user input is the work in [61], which proposes a diversity and usage (DU) map to visualize and analyze EA solutions. Metaheuristics such as Grammatical Evolution, Geometric Semantic GP, and four other configuration EAs are used and compared, and their internal workings are shown in the DU maps. The DU map displays a single heatmap that illustrates the genotype diversity and the degree of gene contribution to a certain solution for multiple algorithm trials. The usefulness of this tool is also addressed in the work with an online questionnaire for many practitioners in the field of metaheuristics to assess the usability and the understanding of the configuration and parameters of the used EAs. With the usage of DU maps, it is concluded that they provide distinct additional benefits to the existing software tools accessible to users, practitioners, and academics who want to reveal important characteristics of the EAs they are developing and examining, promoting their acceptance in real-world applications. In [64], a directional Pareto front is proposed, defined as the objective vectors of non-dominated solutions, and indicates variations in the objective value even in the absence of the Pareto front. Solution representations and the constructed response surface for each tested EA are also present in this work, allowing a better understanding of the process of these algorithms and how different objectives affect solution dispersion, allowing classification of these solutions being on the Pareto front, the supplemental or directional Pareto fronts. This classification via visualization of the different Pareto fronts allows for better DM in the multi-objective optimization landscape.

Recently, visualization tools that intend to interactively show the metaheuristic search process are proposed in [65, 66]. The first tool, called EvoPanel, can effectively and adaptably depict the intricate processes of several metaheuristic optimization algorithms in solving numerical benchmark func-

tions. This mechanism also compares metaheuristic search processes across multiple generations, showing the process of discovering a good solution. The authors even implemented multiple cases where professors, researchers, and students used this tool, showing the general usability of this implementation via a user-friendly graphical user interface (GUI), allowing for interaction capabilities and easier acceptance of such techniques in real-world applications. The second framework implemented uses interactive visualization methods to compare and analyze multiple MOEAs to solve multi-criteria DM problems via scatter plot designs. It compares quality measures, iterations, and solutions of different MOEAs, showing visually three interactive modules: algorithm-level comparison, evolution-level exploration, and solution-level inspection. Similar to previous works, the work implements an interview by expertly evaluating the framework. It highlights its usefulness and general implementation in real-world problems by allowing the implementation of multiple MOEAs and other many-objective problems. Compared to the first tool analyzed, this work needs to consider non-expert analysis and applicability, which should be considered in future improvements to make the work user-friendly and to perceive the complexity of the explainable visualizations.

4.6. Other explainability methods

Other methods that attempt to explain the processes and performance of metaheuristic optimization that do not fit the taxonomy of Table 2 are also considered. These works attempt to provide a theoretical and mathematical analysis of the search procedures of these algorithms, among other multidisciplinary metaheuristic performance explanations.

The 'building block hypothesis' [70] is a prevalent interpretation of how GAs evaluate individuals in a population of candidate solutions. Still, [71] argues that this hypothesis does not explain the evolutionary processes in EAs but rather a consequence of each function optimizer progressively nearing the optimum. Instead, it proposes evolution strategies (ES) as an alternative, emphasizing the collective events occurring in populations that undergo recombination.

The first hitting time and expected first hitting time represent the first time in a single run and the average time for a metaheuristic to find an optimal solution [72]. These metrics describe the average computational time complexity and convergence rate of EAs, offering meaningful insights into these algorithms and their behavior. As such, [72] proposes non-homogeneous

Markov chains to prove problem hardness, i.e., it can only be solved in exponential time. It is proved that for EAs, since they are general optimization methods (problem independent), when their hyperparameters are fixed, the algorithm may progress in the wrong direction, rendering such problems challenging. So, considering this statement a problem is hard if it has an exponentially large search space, which allows the optimization algorithm to go in multiple directions if the exploration and exploitation mechanisms are not well-balanced and fine-tuned.

Local and dynamic optima studies are presented in [73, 74]. Moreover, an undesired property in PSO, stagnation, where convergence to a local optima does not occur, is investigated in [73], where experimental and theoretical fundaments that explain that PSO, even with good hyperparameters almost surely stagnates, that is the expectation is that it converges not to a local optima, but to an undesirable solution. In comparison, and due to the dynamic environments often found in real-world optimization problems, [74] shows theoretical mathematical explanations to monitor dynamic optima and how large populations can be essential in evolutionary dynamic optimization by delineating a correlation between population size and the likelihood of the algorithm deviating from the optimal solution. The authors conclude that population-based metaheuristics are more suitable for these problems than single-solution algorithms, where diversity is a crucial parameter for resiliency and dynamic changes in the objective function. Talking about real-world applications, [75] focuses on improving swarm robotics by developing human-understandable controllers using behavioral metrics and trees. It improves on the conventional swarm management techniques that employ offline EA, frequently resulting in black-box behaviors. This application can enhance safety and resiliency in real-world applications since it is understandable and interpretable by human operators.

EAs based on Non-Darwinian theories of evolution are described in [76] by showing a Peirce evolutionary theory for EAs, which the authors state is superior to the one given by Charles Darwin due to it being more general and more explainable. The semiosis of this theory occurs not just among isolated individuals but is a communal action engaged in by members within a developmental network whose collective judgment finally aligns with the truth, meaning the more general approach taken by Pierce compared to the Darwin theory of evolution.

Explanations regarding metaheuristic convergence, that is, how they arrive at a solution, can be important to understand their behavior and search

process. This mechanism is explained and analyzed in [77, 78]. Explanations regarding convergence analysis of a self-adaptive reduced variable neighborhood search are given in [77] through Neighborhood Guided Evolution Strategies applied to combinatorial optimization problems such as the VRP (NP-hard problem). It examines how simplifying neighborhood operator (swaps, shift, 2-opt, and exchange moves) selection might occasionally enhance convergence in optimization problem resolution. This simplification aligns with the "Less is More" principle proposed in the paper, indicating that minimizing complexity can lead to accelerated convergence and efficient solutions, particularly in large-scale combinatorial optimization problems. The convergence behavior of many-objective EAs is analyzed in [78]. Specifically, it deals with the theoretical aspects of decomposition methods (Euclidean and Tchebyschev norms) and their distance from the Pareto-set, which deal with many-objective problems since Pareto-dominance becomes ineffective when the number of objectives exceeds four. Decomposition methods in EAs formulate several single-objective optimization problems concurrently addressed through a single population's evolution. As such, the work shows insights into how these decomposition methods reach convergence loss (Tchebyschev) and asymptotic convergence (Euclidean) for high-dimensional optimization tasks.

Performance2vec is a framework proposed in [79] to explain metaheuristic performance by generating a vector representation of an algorithm's performance by embedding it within a performance space defined by a collection of benchmark problems. It uses representational learning or embedding methods, where statistical significance and metaheuristic similarities (hierarchical clustering) provide insights into the relations between the metaheuristics and the benchmark problems.

Hybrid approaches, such as the ones in [80, 81], combine EAs with other approaches to improve EA performance and understanding. In [80], a hybrid approach combing evolutionary polynomial regression with multi-objective DE and GA, named EPR-MODEGA, to investigate the search space for models that balance goodness of fit with model complexity. It uses Monte Carlo simulations, which offer an efficient method to measure the average trajectory of model predictions and their associated uncertainty intervals, to measure the influence of uncertainty sources in explanatory data on the simulated output. Sensitivity analysis is also integrated to explore the response of the explanatory variables to improve model robustness and accuracy. In summary, the authors advance the ability of evolutionary algorithms to gen-

erate models with high predictive accuracy, providing insights into optimizing these algorithms for resilient and high-fidelity model development for real-world applications. A multifactorial EA based on diffusion gradient descent (MFEA-DGD) is a method proposed in [81], which proves population convergence and provides theoretical explanations regarding the performance in evolutionary multitasking by facilitating better knowledge transfer across tasks through the DGD implementation through a task convexity measure. In detail, it proposes novel interpretable mutation and crossover operators derived from the DGD's convergence analysis. Incorporating gradient information inside the evolutionary framework as a local search strategy presents an innovative hybrid methodology, advancing the state-of-the-art in multitasking EAs with mathematical fundamentals.

Theoretical explanations for maximizing monotone submodular functions in multi-objective optimization are presented in [82, 83]. Such functions describe the relations between a set of inputs and an output informally, with their importance being associated with several domains, including ML and AI, combinatorial optimization, metaheuristic algorithms, and others due to their characteristics of diminishing returns, where more of one input adds no additional value. As such, maximizing these functions is important, and the works in [82, 83] use EAs for this purpose, providing theoretical fundamentals of the good performance of these metaheuristics on this maximization problem. Specifically, [82] focuses on providing explanations of MOEAs for submodular and approximate submodular functions, i.e., submodular functions with/without a size constraint and maximizing monotone approximately submodular functions with a size constraint. They prove that a straightforward MOEA can attain favorable approximation guarantees for each relevant problem class within the polynomial expected running time. The work in [83] expands on this work and assumes maximizing submodular functions over sequences and not sets as originally used. They explain an MOEA method and why they can consistently achieve or enhance the optimal approximation guarantee after executing in polynomial time on average for these functions over sequence-based problems, such as information gain, recommendation systems, and accomplishing tasks.

Theoretical and mathematical explanations can be fundamental in implementing metaheuristic optimization algorithms in real-world engineering applications. Still, for non-experts and non-developers, these mathematical explanations, even though they provide a foundation and perhaps trust for the end-user, who may not understand anything regarding metaheuristic op-

timization and their mechanisms, these mathematical expressions may even be more complex and troublesome to analyze. So, other methods that are easier to understand and visualize need to be implemented.

5. Understanding the optimization problem and not the metaheuristic

Another way to explain these metaheuristic methods is to understand not the algorithm and its processes but the optimization problem and its influences on the algorithms through their characteristics depending on the application domain. Table 4 shows the taxonomy of the methods used to provide insights into the multitude of optimization problems and their influence on the CI optimization models performance used.

Table 4: Taxonomy of the explainability methods regarding the optimization problem.

Method	**Metaheuristic category**	**No. pubs**	**References**
Landscape analysis	EAs	13	[84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 32]
	Human-based	4	[89, 96, 97, 98]
	SI	1	[91]
Quality-diversity	Quality-diversity optimization	8	[99, 100, 101, 102, 103, 104, 59, 105]
Variable importance	EAs	8	[106, 107, 108, 109, 110, 111, 112, 113]
	SI	2	[106, 114]

EAs: evolutionary algorithms; SI: Swarm intelligence.

5.1. Landscape analysis

Let us denote a fitness landscape, which is composed of three elements (S, V, f), where S represents our solution set, $V : S \to 2^S$ our neighborhood function, and $f : S \to \mathbb{R}$ is the fitness function. These elements can describe a fitness landscape or other type of landscape depending on the definition of f. So, landscape analysis gives information regarding the fitness features that most influence algorithm behavior and performance, which helps understand the problem at hand so algorithm selection and configuration can

be made. In [14], many fitness landscapes, such as multi-objective and error landscapes, are given that go beyond the traditional fitness landscape, and many landscape analysis techniques are also described. The authors differentiate between the following landscapes:

- Multi-objective fitness landscapes: fitness is determined by a multi-objective quality measure, the neighborhood is defined using set-level operators, and solutions are solution-sets;
- Violation landscapes: fitness function replaced by violation function to measure solution deviation from problem constraints;
- Dynamic and coupled fitness landscapes: landscapes where fitness changes overtime due to problems' dynamic environment and coupled landscapes that influence each other in the case of coevolution (objective and subjective landscapes);
- Error landscapes: landscapes mostly used to analyze the optimization process of weight selection in neural network training.

The error landscape can also be referred to as a loss landscape [115, 116], where they are used to understand the connection between learning parameters and model quality in a neural network. In this instance, the landscape serves as an explanation for the ML process itself, as opposed to providing insight into a single model that has already been built, whereas in the position paper in [116], the authors propose the use of landscape analysis not for metaheuristic optimization but to understand XAI methods.

To analyze landscape features and the multiple differentiated landscapes, [14] also discusses multiple landscape analysis techniques such as:

- Bag of local landscape features [117]: analyses fitness patterns in a local neighborhood by calculating binary sequences and the number of better neighbors for a solution vector;
- Codynamic landscape measures [118]: calculates the degree to which objective and subjective landscapes are comparable using measurements of the discrepancies between the landscapes and a sample of points in the search space;

- Constrained landscape metrics [119]: gives five different metrics associated with quantified fitness and constraint violation for a sequence of solutions based on a progressive random walk;
- Degree of separability [120]: uses covariance matrix adaptation evolution strategy (CMA-ES), where nonseparability is measured by evolving the covariance matrix and computing an index from 0 (complete separability) and 1 (complete nonseparability) using the Pearson correlation matrix;
- Exploratory landscape analysis (ELA) [121]: defines six low-level feature classes based on Latin Hypercube sampling;
- Length scale distribution [122]: produces a length scale distribution plot and a single distribution entropy value by calculating the length scales from a sample of solutions acquired via random Levy walk;
- Local multi-objective landscape features [123]: derives landscape features from sequences of solutions obtained through random and adaptive walks;
- Local optima networks (LON) [124]: shows local optima connectivity in a graph-based abstraction;
- Loss-gradient clouds [125]: gives a loss values sample obtained through progressive gradient walks shown in a two-dimensional scatter plot;
- Maximum entropic epistasis [126]: determines direct relationships between variables by computing the maximal information coefficient between each pair of decision variables;
- Population evolvability metrics [91]: measures the evolvability of a population by calculating a value defined as the product of the probability of evolving and the evolutionary ability (ranges from 0 to +inf).

Considering this analysis, Table 5 summarizes each of the identified works concerning the objective landscape, benchmark problems, and analysis technique, putting forth other techniques that [14] did not consider, such as estimating the degree of neutrality and distance metrics and fitness distance, into perspective, among others.

Table 5: Landscape type, benchmark functions, and analysis methods.

Landscape	Benchmark	Technique	Ref.
SO	COCO - BBOB functions	ELA	[32], [92]-[94], [127]
	Quadratic Assignment Problem	LONs	[97, 98]
	NK model	LONs and Coarse-Grained Barrier Trees	[96]
	Sphere, Rastrigin and Schwefel	Extrema graphs	[116]
	Evolutionary robotics	Estimating the Degree of Neutrality	[84]
	Capacitated vehicle routing	Distance metrics and fitness-distance	[85, 89]
	Parallel machine scheduling	Distance metrics	[88]
	Constrained spanning tree	Fitness distance	[87]
	CEC 2013 functions and design of spread-spectrum radar polyphase codes	Population Evolvability Metrics	[91]
	NK and Rought Mt. Fuji models	Adaptive walk analysis	[90]
	Noisy ridge functions	Noise influence analysis	[86]
	River crossing tasks	Movement cost analysis	[95]
MO	ρmnk-model	ELA	[128]

SO: single-objective; MO: multi-objective

The ELA technique is used to characterize the problem features of the COCO benchmark, specifically the BBOB functions in [92, 93, 127, 94] and [32] as Table 3 shows. In [92, 93], ELA identifies pivotal features influencing DE and CMA-ES performance for the COCO benchmark problems, considering ten different configurations for DE and six different configurations for CMA-ES. SHAP values provide global explanations, showing the most essential ELA features. Through ELA, the work in [127] shows that a small number of features, usually less than four, may achieve 98% accuracy in classifying BBOB test functions, minimizing the need for hundreds of features usually needed for configuration models and algorithm selection. A modular CMA-ES framework uses ELA in [94] to evaluate landscape characteristics that contribute to optimal metaheuristic performance regression models. According to the study, feature selection does not always need to be modified for certain settings because the most important features for accurate regression are independent of module configurations. A tool in [32] previously described in other sections describes intrinsic algorithm explainability and uses ELA features of the BBOB benchmark functions to evaluate algorithm performance based on those features. Random forests (RF) and simple implementations of decision trees (DTs) are used to train multi-output models for algorithm performance forecast and to understand which type of landscape and its features are important in selecting a configuration or module for the tested metaheuristics.

Other works use LONs for landscape analysis regarding metaheuristic design, [97] uses machine learning techniques to predict multiple LON algorithm features that impact the search performance of heuristics in a quadratic assignment problem. The authors used iterated local search and TS as response variables for the predictive models, where RF achieved a higher success rate to explain these algorithms' variance. Similarly, in [98], the authors analyze the LON funnel depth and how it influences different acceptance criteria for the iterated local search algorithm in combinatorial optimization. The work outcomes found that search difficulty and funnel depth are related and that acceptance technique and funnel shape interact, where up to 96% of the variance in ILS duration might be explained by regression models using funnel depth data.

Two other manuscripts, taking LONs as a basis, proposed other landscape analysis methods, such as grained barrier trees [96] and extrema graphs [116]. The first presents a coarsely-grained tree for multiple instances of the NK model, which clusters LONs for better visualization and characterization in

fitness landscapes. In this work, a novel flooding algorithm variation that works with LONs is the main component in the barrier trees. The outcomes of this research suggest that the tree depth may impact how challenging the landscapes are to explore in an iterated local search. One possibility might be that impediments keep iterated local search iterations from breaking out of local optima clusters. The second, an extrema graph visualization approach of the local structure around a potential global optimum, is proposed, and it was first introduced for isosurface extraction in volume visualization. This method enables understanding the search space because these maps incorporate global information detailing the spatial interactions between extrema. This implementation is intended for metaheuristic experts to comprehend the process and performance of these algorithms.

One of the obtained works in this review analyzes the estimated degree of neutrality in fitness landscapes through a standard genetic distance method between populations [84]. The authors applied such methods to a real-world optimization problem involving the realization of tasks in robotics, where the investigation of the transitions in the genetic distance between populations over a certain number of iterations needs to happen to confirm the existence of neutrality in such a problem. With this application, an assumption can be made of the evolutionary processes for problems for which traditional fitness landscape formulations can be explained. Multiple other works also use some distance metrics and fitness distance between solutions to analyze fitness landscapes [85, 89, 88]. Fitness-distance analyses are performed for capacitated vehicle routing problems in [85] and [89], where the first focuses on potential distance variations between sets of local optima and random solutions of a specific instance and also attempts to find trends in local optima solutions. In comparison, the second work goes further and also analyzes the influence of the neighborhood operator of a metaheuristic for two different landscapes, where the performance of these operators is analyzed through the autocorrelation of the fitnesses of both landscapes. These works aim to provide insights into metaheuristic performance, mainly local search methods, in combinatorial optimization and explain why they performed well with different operator selections. Similarly, [88] proposes a weighted distance metric for a parallel machine scheduling problem, where a single-objective version of the Non-dominated Sorting Genetic Algorithm (NSGA-II) is applied. In this case, the correlation between the distance to the optimum and the objective function value is analyzed, obtaining a smooth and monotonic correlation implying that the algorithm will extract relevant information for

the global search from a local sampling of the objective function when such a weighted distance directs it. Correspondingly, the work in [87] presents the same fitness-distance analysis as the previous works for multiple instances of constrained spanning tree problems, using a multitude of solution encodings analyzing its impact on the cost function and the different distance metrics.

In [91], the population evolvability metrics landscape analysis method that was previously described can be found for proper algorithm selection. The authors start from the concept of dynamic landscape analysis, where optimization algorithm information is improved by applying statistical analysis to population-based metaheuristics. Success probability, convergence velocity, and convergence reliability are used and combined to measure population evolvability and algorithm selection tasks according to statistical significance for a real-world problem.

The works in [90, 86] explore the changing geometry of NK fitness landscapes along an adaptive walk and noise usefulness in fitness landscapes, respectively. The work [90] considers two types of landscapes with variations of the K and N parameters, where they concluded that the increasing sign epistasis and decreasing antagonistic-to-synergistic ratio at higher fitness levels, supporting that this phenomenon in adaptive walks is due to mean regression, not inherent properties of the walks. The work [86] contemplates noisy ridge functions using a standard noise model of additive normally distributed noise for ES. Primary to contrary belief, this work shows that including noise improves EA performance, where premature convergence is avoided since the algorithm is no longer at risk of falling in a local optimum. Another work in [95] uses the movement cost of agents' genomes in a river-crossing fitness landscape. It concludes that restricting the search space would provide additional information and bias the problem. The authors even propose a principled simplification to reduce the dimension of the testbed to analyze key problem features to gain agent explainability.

The only work that uses a landscape other than a regular fitness landscape is [128], where a multi-objective fitness landscape is analyzed through ELA, where multi-objective landscape and decomposition-based features are considered for algorithm selection and performance forecast. In more detail, the work identifies distinct global trends in feature values that lead to significant cost-vs-accuracy trade-offs, particularly using random walk with a few neighbor solutions.

Even though landscape analysis is a powerful tool to inform experts about algorithm performance and selection regarding landscape features prediction

that most impact algorithm behavior, this is a very complex field that non-experts cannot understand. Therefore, further work needs to be done so that this landscape analysis's representation can be easily understood and interpretable enough by non-experts.

5.2. Quality diversity

QD optimization is a recent topic in EC, where the main goal is to illuminate the search space, storing the best-performing solutions according to feature descriptors or behavioral characteristics that are interesting to analyze for the problem in question [101]. As such, QD algorithms do not operate in a standard parameter space but in a feature space. Diversity is the most critical factor in QD, and the algorithms try to fill the whole feature space with the best-performing solutions for a performance descriptor.

In this new paradigm, the Multi-dimensional Archive of Phenotypic Elites (MAP-Elites) [99] algorithm stands out. This algorithm can also be described as an illumination algorithm, which uses an archive structure divided into regions to use the best solutions, or elites, based on the problem behaviors or features the user deems relevant to analyze. This structure illuminates the search space by making the user aware of how these features impact metaheuristic performance. This algorithm presents a major drawback, as [102] specifies, where it cannot scale for multiple dimensions of feature spaces because the size of the archive structure increases accordingly. Still, in [102], the authors improve the MAP-Elites algorithm by addressing this drawback. The authors implement a centroidal Voronoi tessellation to divide the feature space into a desired number of regions. It inserts each created individual in the nearest region, substituting a less fit one (through individual performance) if the region is already filled.

Other works, such as [100], attempt to understand the influence of the behaviors in QD algorithms. In this study, the freedom of the evolutionary process from the requirement of defining measures of algorithm diversity and fitness is proposed. It is concluded that the real objective is to identify algorithms that do not need the definition of problem behaviors prior, and care must be taken while comparing algorithms about the calculation of the diversity, as various formulations may contain errors. While in [105], different encodings of voxelized meshes in an architectural scenario of QD optimization (direct encoding, a dictionary-based representation, parametric encoding, compositional pattern-producing networks, and cellular automata). The

findings show that certain indirect encodings can provide more varied solution sets and perform better than direct ones, particularly when considering whole phenotypic variation. Competition between species of encodings based on phenotypic characteristics results in a performance close to the optimal single-encoding QD method, meaning that QD optimization only sometimes needs the top-performing single encoding to contribute.

QD optimization can also be integrated into selecting hyper-heuristics, another topic addressed in this review, as [59] demonstrates. Here, the viability of optimizing a sequence-based selection hyper-heuristic model via MAP-Elites is investigated for combinatorial problems. The usage of QD optimization is to provide information regarding hyper-heuristic sensitivity based on multiple behavior indicators (15-dimension archive).

In comparison, the works in [103, 104] attempt to further provide trustability and explainability to QD optimization, which already presents some intrinsic explanations for metaheuristic performance. In [103], a support technique to aid the user in the DM process is proposed via a parallel coordinate plot with user interaction capabilities. The work presents a visualization of behavior interaction highlighting user-specific criteria, where domain experts without algorithm knowledge will be integrated in future work. In comparison, [104] presents an automatic, explainable rule set extraction from the elite archive that represents the relationships between the behaviors and solution quality for a high-dimensional archive. In this work, user expectations are also considered, which are matched with the extracted rules, involving domain experts in the results obtained by the algorithms, taking a step further in understanding the quality of the solution in a real-world context.

5.3. Variable importance

An interesting topic is to analyze the variables in optimization problems that impact the metaheuristic performance and how they correlate. Table 6 shows the works that analyze the variable impact on metaheuristic behavior.

The analyzed works extensively used Data mining techniques for variable ranking and interdependence. For instance, [106] proposes ParetoGP for variable importance and selection to understand which variables most influence the flavor liking scores instead of other mining approaches such as trajectory mining. On this basis, knowing variable significance enables model similarity grouping and efficient sensitivity analysis. A multi-objective PSO was also implemented to maximize the flavor liking of assessors and group assessors. This work then allows a better DM process since it provides explanatory

Table 6: Methods, benchmarks, and performed analysis regarding variable importance.

Ref.	Year	Method	Benchmark	Analysis
[106]	2012	ParetoGP and variable sensitivity	Flavor optimization	Variable importance and correlation
[107]	2020	Data mining design space exploration	Embedded Systems Synthesis Benchmarks Suite	Mapping variable pattern lengths
[108]	2021	PCA	1D Checkerboard, Royal Road, Trap-5	Information gain and PC loadings
[109]	2021	Surrogate mining	BinVal, AltOnes, Checkerboard	Variable ranking scores
[110]	2021	Surrogate mining	Cellular windows	Variable combinations for local solution sensitivity
[111]	2022	Surrogate mining	1D and 2D Checkerboard, Trap-5, MAXSAT	Variable contributions for metaheuristic generations
[114]	2023	Movement of variables	Bidding in local electricity markets	Variable representation and ranking
[112]	2023	PCA and surrogate mining	1D Checkerboard, Royal Road, Trap-5	Information gain and correlation of PC loadings and variables contributions
[113]	2023	Search trajectory and variable sensitivity	Rostering problem	Variable convergence and ranking

capabilities, enabling the integration of protocols in this area. In [107], automatic data-mining methods are incorporated into Design Space Exploration (DSE) frameworks when employing EAs for embedded systems benchmarks optimization. Data mining is employed to automatically extract and generate previously unexplored information from the optimization process, which is then integrated into the DSE to improve the solution quality and provide further explainability for the EA. Interdependence between decision variables and optimization objectives by ranking high- and low-quality solutions is demonstrated during DSE by mapping variables, given by pattern lengths, which shows their influence regarding solution quality and objectives.

The works in [108, 112] use Principal Components Analysis (PCA) trajectory mining for dimensionality reduction to extract key problem features and facilitate visual analysis of metaheuristic behavior [129]. Both studies measure local and global information gain by analyzing benchmark problems (1D Checkerboard, Royal Road, Trap-5) solved by GA and a modified population-based incremental learning method. The latter is compared to the angle from the origin to track knowledge obtained. By recording patterns over time, the work [108] illustrates how dynamic variables contribute to fitness by showing PC loadings every generation for every method and challenge. In [112], variable contributions to fitness are visualized by mining surrogate fitness models of final solutions. Spearman rank comparisons demonstrate positive rank values for each problem, and metaheuristics validate the close alignment between surrogate modeling results and PC loading findings. These approaches provide fundamental explanations by relating changes in information gain to varied importance rankings and interaction features. The results of surrogate mining and PCA concur, indicating that the underlying dynamics of the problem, not the explanation methods, are the source of these explanations.

In comparison, [109, 110, 111] use data mining of surrogate models to provide problem variable importance. Multiple ML surrogate models, namely RF, DT, and linear regression, were employed in [109] to provide a variable ranking score that indicated each model's significance for several binary benchmark problems (BinVal, AltOnes, and Checkerboard). These ML methods were combined with a GA, and the variable ranking indicated which variables may be dropped for the given problems if they were not needed. The authors in [111] propose an extremely similar approach using a Support Vector Regression (SVR) model to generate the surrogate models of four benchmark problems (1D and 2D Checkerboard, Trap-5, and MAXSAT)

using population data provided by a GA. They analyze and compare the knowledge acquired by the surrogate model after being trained on population data from the first generation with that of a surrogate model trained on population data from all generations by representing the variable contributions. It is advisable to train the surrogate on all solutions encountered throughout the execution of the GA instead of doing it in the first iteration of the GA. Meanwhile, [110] uses an approach for comprehending the metaheuristic choice process by showing which variable combinations affect solution quality. The authors extract solutions from a surrogate fitness model (Markov network fitness model) and examine the local sensitivity of those solutions concerning the mean coefficient energy values and the least cost for the facade of a building. This sensitivity aids the decision maker in trusting metaheuristic results if the solutions obtained align with the conclusions derived from the models' formulations.

The work in [114] takes inspiration from the previous works and considers a bidding problem in local electricity markets using an ACO for the optimization process, where a representation is done regarding the solution encoding for the problem understandably so non-experts and other users can understand the problem and metaheuristic process at hand. A post-optimization analysis is done, using variable movements by applying a perturbation with a 1% disturbance, and the fitness is recorded so a variable ranking and fitness disturbance can be generated. The authors concluded that variables associated with combined heat and power generation offered in the local market are the most important since they affect fitness the most. In [113], an approach to extracting relevant explanatory features by analyzing the search trajectory of a GA via PCA is implemented for a staff rostering problem. A post-optimization analysis is made, where the outcomes are put through a sensitivity analysis and further compared with the trajectory analysis. Throughout the search, the methods uncover the extent to which individual workers impact the overall quality of the rostering solution and facilitate the identification of workers who have a lesser influence on the overall fitness of the solution and those who have a higher flexibility.

These last two analyzed works consider real-world applications when applying these explainability methods compared with the other works in this section that only consider standard academic benchmarks. Still, these implementations are considered general since they explore the problem and its key features other than the utilized metaheuristic algorithm, so they are general enough that other algorithms can be implemented with easier applicability

to other problems. Still, non-expert user input should be considered in their implementation to identify whether these methods are easily understandable or extremely complex for non-developers and non-expert users.

6. Shedding light on machine learning models using metaheuristic algorithms

Metaheuristic algorithms can be used in a hybrid fashion with ML methods, where they optimize counterfactual explanations, classifier systems, knowledge graphs, and explainable recommendations so ML forecasts can be interpretable. Genetic programming is an EA with interpretable XAI concepts behind its evolutionary processes, so the intrinsic processes of this algorithm are already explainable to EA experts. Table 6 demonstrates the taxonomy of the methods that use metaheuristic optimization to provide understanding for ML models.

Table 7: Taxonomy of the explainability methods for metaheuristic optimization for ML models.

Method	**Category**	**No. pubs.**	**Ref.**
Counterfactual explanations	EAs	5	[130, 67, 131, 132, 16]
	SI	1	[133]
Genetic programming	EAs	17	[134, 135, 106, 136, 17, 137, 61, 19, 138**?** , 28, 168, 104, 169, 167, 171, 170]
Knowledge graphs and explainable recommendations	EAs	4	[139, 140, 141, 142]
Rule and recommender systems	EAs	26	[134, 135, 143, 144, 145, 146, 147, 136, 17, 148, 149, 150, 104, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163]
	Human-based	1	[151]
	Physics-based	3	[136, 164, 165]
	SI	2	[145, 136]

EAs: evolutionary algorithms; SI: swarm intelligence.

6.1. Counterfactual explanations

In XAI, explanations that potentially answer "what-if" scenarios when using ML methods can be denoted by counterfactuals. These explanations are

intended to offer users all the possible cases, even the worst ones (if they fail), regarding the models' performance so trust and understanding can be gained in the model. Metaheuristic algorithms can generate these explanations by constructing all these possible scenarios, as in [130], where an approach generates counterfactual explanations for a Deep Learning model through an adversarial strategy. This adversarial strategy is plausible, where modifications applied to the model's input provide a semblance of legitimacy and believability without computer involvement. The authors employ a multi-objective metaheuristic framework to develop counterfactual explanations for a given image, comparing NSGA-II, Generalized Differential Evolution, and Speed-constrained Multi-objective Particle Swarm Optimization, where the latter obtained a better performance in generating these counterfactuals. These explanations aim to balance three objectives associated with the classifier's explainability: plausibility, adversarial success, and intensity of the attribute perturbation.

The works in [131, 132] introduce model-agnostic counterfactual explanations generation methods. The work in [131] presents an evolutionary counterfactual model for time series classification based on NSGA-II. The proposed framework is evaluated on uni- and multivariate time series, where it outperformed other algorithms, generating correct counterfactuals by only changing a small portion of features. The algorithm also outperforms others for metrics such as proximity (mean absolute error) and sparsity (fewer changes from baseline), and on plausibility, it performed similarly to the other algorithms. In contrast, [132] proposes a novel algorithm named CounterFactuals Now (CFNOW) based on TS to efficiently generate counterfactuals for several data types (tabular, picture, and text) using a modular and rapid two-step procedure. They verified the efficiency and applicability of CFNOW using similar metrics: coverage, sparsity, distance, and generation time, which were compared with other counterfactual generation algorithms for different dataset types, and it was able to outperform the compared algorithms by generating more valid explanations (when the features change) in less computation time and effort.

In a similar fashion, [16] proposes CREATED: CounteRfactual Sequence generation with Evolutionary AlgoriThms on Event Data, a general framework for counterfactuals generation, which does not require any knowledge of the environment that the ML model is being used. It involves training a Markov model to determine the viability and feasibility of creating counterfactual sequences considering the outcome prediction, similarity, and

sparsity. The results demonstrate that this method outperforms baseline algorithms such as case-based, sample-based, and random baselines regarding viability by generating more valid high-quality counterfactuals without considering the problem domain, making this approach highly applicable to any ML model in any situation.

The work in [133] improves from the previous works, and it uses Niching PSO to generate multiple sets of counterfactual explanations, named Roshomon sets (full sets of optimal explanations), using different data features. It compares and examines different Niching approaches and assesses the most effective approach for generating these sets. The best Niching approaches to generate the counterfactual explanation sets are a Fitness Euclidean-distance Ratio, and Ring Topology PSO, where the Ring Topology is preferable when seeking a more varied Rashomon set, and the Fitness Euclidean-distance Ratio PSO should be employed when aiming for the highest possible overall fidelity of the Rashomon set. The application of these methods in a real-world scenario was also examined, demonstrating that the algorithms can be seamlessly incorporated into an existing ML ecosystem with minimal adjustments due to the small number of hyperparameter tuning needed as well as respecting all of the optimal explanation subtopics: contrastive, selected, social, and do not only focus on probabilities.

Human integration for counterfactual explanations is integrated in [67] for fuzzy classification systems. The authors thoroughly examine qualitative explanation-generating techniques that utilize language concepts and a quantitative method via GA to modify specific feature values. Through online questionnaires, two human assessment studies analyze these approaches using a perceived explanation complexity score. The findings indicate that, although they require more work from participants, within-subject surveys with complex frameworks offer deeper insights into the quality of explanations than in simpler settings. This work emphasizes the value of user-centered methods in explanation systems and enhances XAI by integrating human components, particularly for those not specialists in AI.

6.2. Genetic programming

In the world of metaheuristics, GP is an EA where each individual (solution) of the population is a computer program or a function used to determine a known target considering specific input values. So, GP acts more as an ML method, where an optimal program or function needs to be determined while optimizing the error to the target values [166]. This program is given in a

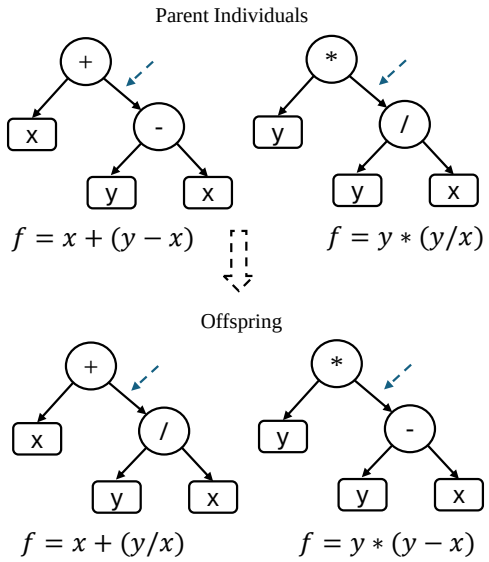

Figure 7: Individual and one-point crossover GP tree representation, adapted from [18].

tree-like structure where the user can interpret every branch as Fig. 7 shows. In this case, the figure demonstrates a one-point crossover to control model size. The subtree $y-x$ from parent one and y/x from parent two are swapped, as they inherit the same structure as the upper part of the parents [18].

As such, GP is an explainable method for ML mechanisms since it presents a good balance between explainability and model accuracy, as Fig. 8 demonstrates [13]. The survey work in [18] distinguishes using GP for intrinsic interpretability, where GP works directly on providing explanations on the one hand and using GP in post-hoc interpretability to elucidate other black-box ML models and visualization methods for and by GP. In this taxonomy, the authors give the name of XAI by GP (XGP).

Table 8 shows the application domain and the taxonomy regarding the interpretability in GP of the obtained works incorporating GP. For instance, the works in [134, 135, 136, 137] show the application of GP for water resources and quality management, as demonstrated in the previous section. The first and second works use GP to evolve rule set structures for nonlinear relationships between physical and chemical variables. The third work also uses GP to evolve explanatory rule set models optimized by different metaheuristics as described in section 6.4 to forecast not only the concentration of chlorophyll-a but also the concentrations of the potentially toxic cyanobac-

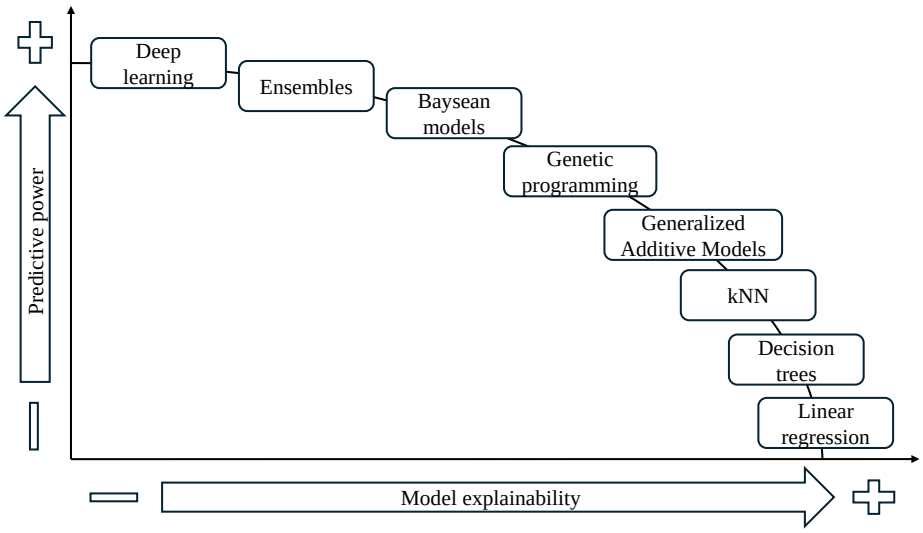

Figure 8: Model explainability and predictive power trade-off for some ML models, adapted from [13].

terium in Queensland, Australia. The final work in this subject proposes a short-term forecast model by implementing GP. GP is employed with a hypothetical rainfall-runoff model to address the model setup problem of two catchments in Singapore and New Zealand, showing its good compromise between accuracy and complexity. In all these works, GP directly explains the problems at hand, corresponding to intrinsic interpretability.

The usage of GP for product development is described in [106]. In this work, ParetoGP is implemented for intrinsic explainability by extracting and analyzing explanatory variables to construct all feasible models that elucidate the provided data intended to understand human behavior in product development. In [17], a rule-based hybrid approach of an extension of GP named Genetic Network Programming (GNP) with an NN is proposed to extract stock trading rules combined with technical indexes. This work uses GP to inherently generate interpretable models via rule sets in stock trading predictions. Moving into the medical field, the work in [138] proposes a cartesian GP (CGP) approach for automatically detecting Parkinson's Disease through patients' handwriting and health protocols. Here, the cartesian GP facilitates the assessment of the efficacy of the most informative problem features and offers a clear model of classification criteria through its intrinsic interpretable structure, which can be utilized as a framework for developing

more effective evaluation tests compared to other ML models, such as RF, SVM and others. The authors in [104] analyzed previously in other sections, use GP as a classifier to generate interpretable rule automatically sets from a MAP-Elites archive on an urban logistics problem, with AND, OR, and NOT connectors, so they can be interpreted and evaluated through expert users belief system. In [167], a co-evolutionary process approach combines CMA-ES and GP to optimize vision and decision modules (glass-box models) for reinforcement learning (RL) on multiple Atari video games. Specifically, GP is used to evolve DTs, inherently contributing to the interpretation of these structures, that is, in the construction of the solutions of the RL itself. This approach ensures inherent transparency in the DM process rather than relying on typical post-hoc explainability methods, which may not accurately represent the model's genuine DM process.

Some works use GP for XAI (XGP) interpretability as [19, 168] demonstrates. Multi-run subtree encapsulation to provide explainable solutions in XAI is proposed in [19]. In this work, if the solution obtained contains encapsulated subtrees (simple algebra expressions), these are computationally straightforward, and, in theory, the overall expression can be computed efficiently and provides explainability of the generated formulations. As such, the intrinsic explanations provided by GP can be interpreted and even elucidate the mechanisms underlying black-box methodologies reliant on ANN, including DL (post-hoc interpretability). The [168] proposes a model learning approach with personalized interpretability estimation, ML-PIE, to consider user feedback by assessing their preferences during the GP evolutionary process. It shows the users the obtained models and asks which is more interpretable, adjusting the ANN estimator for explainability and internally giving the most interpretable model for the GP process. GP Gene-pool Optimal Mixing Evolutionary Algorithm (GP-GOMEA) finds small expressions in GP, so these can be interpretable [18]. Still, this mechanism cannot reuse subexpressions, which can be mitigated by using CGP, which [169] proposes by the name of CGP-GOMEA. It uses a predetermined quantity of rows and columns of nodes in the GP structure. With CGP-GOMEA, it is possible to find models of smaller sizes than the standard GP-GOMEA, enabling a better intrinsic understanding of the provided expressions.

The works in [61, 170] are multidisciplinary, without a specific application domain. In [61], visual representations of the GP evolutionary process are used. It uses DU maps to gain insights into the evolutionary process of EAs, namely GP, and understand their tree-based structures and diversity

mechanisms. A context-aware evolutionary learning algorithm (CELA) is introduced in [170] to reduce computational effort and increase interpretability. CELA creates explainable models with many features for large datasets using GP-based symbolic regression and domain expertise.It improves the evolutionary feature synthesis approach by offering better accuracy while maintaining interpretability, outperforming black-box models, namely DL and ANNs, and allowing a better DM process.

Table 8: GP taxonomy and application domain for ML.

Application domain	**Taxonomy**	**Ref.**
Water quality management	Intrinsic	[134, 135, 136, 137]
Product development	Intrinsic	[106]
Stock trading	Intrinsic	[17]
Pathology's diagnosis	Intrinsic	[138]
Urban logistics	Intrinsic	[104]
Video games	Intrinsic	[167]
XAI	Intrinsic and Post-hoc	[19]
	Intrinsic	[169, 168]
	Intrinsic	[170, 171]
Multidisciplinary	Visual	[61]

The influence of genotype-phenotype mapping structures on the evolutionary search mechanism in GP is analyzed [28] on complex networks. This work is concerned with providing explanations of the internal dynamics of GP rather than how GP could be used for model interpretation or explanation. It provides researchers with more expert search mechanisms that can make GP and other EAs more robust (the ability of a phenotype to resist mutations) and more efficient across different problem instances.

Periodic boundary conditions have enhanced GP's robustness and performance for the CEC benchmark functions without real-world applications [**?**]. This mechanism focuses on enhancing GP performance for optimization, emphasizing the exploration capacities of GP to prevent premature convergence to local optima, a prevalent issue in unimodal optimization. Still, the structure of GP allows an understanding of experts of the evolutionary processes due to its program structure. Population diversity and genetic inheritance can be important in enhancing and understanding GP performance [171]. Diversity is important in avoiding premature convergence in the search space to suboptimal solutions. By integrating genotype and phenotype measurements, [171] elucidates diversity loss and distinct traits of the

evolutionary search of GP and other EAs. Still, this evolutionary search in most metaheuristics is significantly influenced by their parameterization and problem features.

6.3. Knowledge graphs and explainable recommendations

Visual structure representations of information, capturing relationships between multiple entities, are denoted as knowledge graphs. These graphs enable a more meaningful data analysis of those systems, allowing for a better decision-making process. It establishes meaningful connections between data points, enhancing comprehension of complex information and assisting in activities such as recommendation systems, semantic search, and AI-generated insights in many fields. In [139, 140], explainable recommendation systems are proposed based on knowledge graphs, which connect users and items, and the path is given as an explanation to the target user regarding the provided recommendation. The first referenced work compares multiple MOEAs to optimize three objectives: accuracy, diversity, and explainability of the recommendations on two datasets. It also constructs the knowledge graph based on semantic relationships to establish connections between users and items. The explainability of a suggested item is measured by the product of the user's rating of the item with the similarity of attributes across items. The authors show that their proposed framework outperforms other algorithms for the three considered objectives with higher explainability in the generated recommendations, illustrating the explanation paths in the built knowledge graphs. Similarly, the second work compares many-objective EAs with four different objectives, improving from the previous work by adding the novelty objective. In this case, the knowledge graph is constructed using triples generated by identifying entities and relationships. The significance of the relation to the entity is reflected by calculating the product of the corresponding vector of the entity and its relation to the items, which provides the interpretability of the recommendation. The outcomes are similar to the previously analyzed work, where the proposed algorithm outperforms other state-of-the-art methods, and a visual representation is given via a knowledge graph with explanation paths for a user and the chosen item.

Ontologies are often used to define relationships and entity types inside knowledge graphs. The ontologies can be mapped by concepts in description logics. Still, in most cases, the models used for learning description logics are black-box models, not allowing the learned model to be enriched with domain knowledge. EvoLearner [141] is a tool used to tackle these challenges,

using EAs to efficiently and understandably learn these description logic concepts, where syntax trees form the individuals of the EAs. As such, this tool can deduce new relationships or corroborate existing ones, increasing the structure and reasoning capabilities of knowledge graph construction and refinement.

One other work, which is a proof-of-concept, presents Fuzzy Cognitive Maps (FCM) based on knowledge graphs for analyzing operational environments where a single action or a series of actions result in a chain of effects that significantly changes important end-chain consequences [142]. It uses a GA to optimize all knowledge graph solutions, where solution representation is compared with the fitness function based on the multiple effects considered (targeted, neutral, and undesired). These effects allow the GA to find the optimal region of solutions. The authors state that knowledge graphs or FCM systems offer a viable and effective method for capturing the intricacies of effects-based operations environments while maintaining resilience and explainability.

A knowledge graph already explains the DM process since it represents the connections between entities with the provided information. Still, user preferences and inputs need to be considered when designing such models so they can understand the complexities associated with these decisions.

6.4. Rule and recommender systems

Rules and recommendations present another interpretability area for optimization problems by evolving conditional model structures and recommender systems involving expert knowledge. Hence, users understand their impact on the problem by providing DM interaction capabilities. Rule-based classifier systems can be created via metaheuristic optimization algorithms to provide insights into ML models and their predictions in a hybrid formulation integrating both metaheuristic algorithms and ML models. Table 9 demonstrates the proposed classifier methods and the application domain of the analyzed works.

Hybrid EAs (HEA) to evolve rule sets of "IF-THEN-ELSE" models in freshwater ecosystems are proposed in [134, 135, 136]. The first and second works only use GA to optimize the parameters of the rule structures. In contrast, the third work evolves the first and second works by using multiple metaheuristic algorithms based on local search, evolutionary, and swarm processes to optimize the parameters of these rule models. In the second work, the authors show the best rule set models obtained by each metaheuristic

Table 9: Methods and application domains for rules and recommender systems in ML domain.

Ref.	Year	Type	Method	Application domain
[134]	2006		HEA	Water quality management
[135]	2007		ANN + HEA	Water quality management
[143]	2007		MILCS	Multidisciplinary
[144]	2008		MILCS	Multidisciplinary
[145]	2010	Rule-based	DIFACONN-miner	Multidisciplinary
[147]	2013		DEREx	Medical
[146]	2013		SVM-CC	Medical
[136]	2014		HEA + PO	Water quality management
[17]	2014		MLP + GNP	Financial
[148]	2016		MOEDA/D	Multidisciplinary
[149]	2017	Recommender	DEMO	Energy
[150]	2017		DEMO	Energy
[104]	2021		GP + CN2	Urban logistics
[151]	2021	Rule-based	RATP	Financial
[152]	2021		MORE	Software development
[157]	2022	Recommender	X-SBR	Software development
[156]	2022		HAUS-rules	Medical
[153]	2022		ARM + EA	Metro map creation
[154]	2022		GARAbDG	Multidisciplinary
[155]	2022		PALM	Financial
[158]	2022		DEREx	Medical
[165]	2023	Rule-based	SA + IT2FLS	Energy
[164]	2023		SFO	Construction
[159]	2023		MRSL	Multidisciplinary
[160]	2023		MOQAR + LSTM	Time series forecast
[161]	2023		DEREx	Medical
[162]	2024		XLEMOO	Multidisciplinary
[163]	2024		XIMEA	Multidisciplinary

with an error comparison. They also present them in a user-friendly manner that field experts can interpret regarding water quality conditions.

Learning Classifier Systems (LCS) are rule-based machine learning models combining EAs with learning methods such as supervised, reinforcement, or unsupervised learning. Mutual Information LCS (MILCS), proposed in [143] and [144], enhances LCS by employing mutual information as a fitness metric, outperforming traditional metrics, namely accuracy. This approach identifies generalizable rules and offers explainable rule-based models with visualization methods. A systematic evaluation framework for explainable rule-based models, including metaheuristic rule set learners (MRSLs) in particular LCS and fuzzy systems, is introduced in [159]. Using synthetic datasets, the XCSF classifier system demonstrated the ability to evaluate algorithms in a controlled environment, improving performance comparisons.

ANNs and other DL models can achieve very high classification accuracy, but as stated, these methods are black boxes, and their explanation power is basically non-existent. The work in [145] tackles this challenge by proposing a DE algorithm for training ANNs and utilizing an ACO technique for rule extraction, designated as DIFACONN-miner. It evaluates ANN fitness according to performance measures such as error, rule number, and accuracy. The comprehensible extracted rules provide insights into the underlying patterns learned by the ANN, which allow domain experts to interpret forecasting results more easily, bridging the gap between model accuracy and transparency and transforming ANNs into white-box models.

In the same spectrum as ANNs, support vector machines (SVMs) are considered black-box models with notable accuracy but reduced explainability. In [146], an explainable two-step hybrid approach in which SVMs conduct learning and a clear emulation of the resulting decision model is produced by cooperative coevolution (CC) in the form of propositional rules in the medical field. This field requires interpretable and extremely accurate models for the DM process. The interpretability of the rules is enhanced via feature selection, facilitating swifter and more precise classification. Still, in the medical field [147, 158, 161] propose a DE-based rule extractor (DEREx) classifier process. Each solution of the DE algorithm contains an IF-THEN rule set with AND-connected clauses intended to extract automatic knowledge from the databases. The described works contribute to the field as they offer significant insights for medical professionals, as relationships among various medical indicators are automatically identified, and their reliability is guaranteed by computational analysis. Furthermore, the individual categorization ex-

planation enables the medical professional to discern a hierarchy of attribute significance in the automated diagnostic process. Another application, in the medical field, mines high average-utility sequential rules (HAUS-rules) using logical operators via MOEAs, where utility and interestingness are maximized [156]. The authors of this work state that the HAUS-rules MOEA is an extension of NSGA-II for evolutionary mining and antecedent condition selection, considering rule length for easier interpretation. This technique enhances standard sequence mining by emphasizing high utility and interpretability sequences. Assessing the rules obtained by obesity-field experts allowed for a better representation, and the proposed tool proved to be a potent instrument for biomedical research, such as analyzing longitudinal gene expression data.

Attribute-based Decision Graphs (AbDGs) are an alternative graphical representation of vector-based data appropriate for multiclass categorization. It presupposes that the data set has p attributes, with each attribute segmented into mutually exclusive intervals of values. The limitation of the AbDGs is their dependence on structural definition, which [154] tackles by proposing an explainable classifier system based on genetically evolved graph structures. It uses a GA for rule attribute-based decision graphs (GARAbDG) for explainable classification. The classifier is constructed to be explicable by associating decision nodes with class probabilities, enabling users to follow the rationale behind the classification.

In [104], GP and CN2 are used for rule extractions from the MAP-Elites archive structure on an urban logistics problem. The GP algorithm uses AND, OR, and NOT connectors for rule generation, and CN2 uses "IF-THEN" conditions for these rules. The work also integrates the domain experts for post-hoc analysis. A novel metaheuristic algorithm with XAI principles is proposed in [164] for concrete compressive strength classification and mixture design of recycled aggregate concrete. The implemented metaheuristic is an adapted sunflower optimization (SFO) method that balances predictive accuracy with rule simplicity, enhancing interpretability for practical applications. The rule-based SFO combines optimization with XAI to improve forecast accuracy and provides interpretable outputs for practical applications in concrete design.

In the finance world, [151] shows a multi-objective hybrid EA approach for credit risk assessment, which the authors named rules acquisition with threshold as particles (RATP). The interpretable classifier is a collection of concise linguistic interpretable rules obtained from the high-dimensional data

without a preprocessing burden. It offers insights for DM, especially in financial applications encompassing discrete and continuous data, frequently characterized by missing values. Regarding stock trading, [17, 155] show ensemble and online learning mechanisms to handle market trends and uncertainty. Specifically, the first work shows a hybrid mechanism combining a rule-based EA with a multi-layer perceptron (MLP) to enhance buy-and-sell DM using a rule pool. Even though the rules generated by the EA are interpretable by experts in themselves, the selection process of the MLP and an explanation as to why a certain rule was selected is a challenge to the full interpretability of the proposed mechanism. A rule-based autonomous neuro-fuzzy learning system, the parsimonious learning machine (PALM), is presented in the other work to predict time-varying stock indices, where model explainability is considered. An implemented type-2 fuzzy logic system and its economic structure ensures reduced complexity while preserving prediction accuracy, which the authors named T2-PALM. MOEAs were applied to optimize model accuracy and explainability via the number of rules (fewer rules = more interpretable), enabling real-time adaptation to fluctuating market conditions. Similarly, but in another field, [165] proposes a metaheuristic-based interval type-2 fuzzy logic system (IT2FLS) for solar radiation modeling, which is a valuable input in photovoltaic (PV) generation forecasting. The SA algorithm performs rule selection and limits the linguistic variables in each rule. So, a short rule list is provided for expert analysis to refine lifelong learning models by combining their experience.

In architecture-based software, [152] proposes a multi-objective performance optimization based on explainable randomized search rules (MORE). This approach balances performance with interpretability by simplifying rule-based solutions, where the performance and explainability of such a framework are compared to other methods, metaheuristic and rule-based, showing the higher statistical significance of the MORE tool. Still, close work with professionals and experts in addressed fields can and should be considered with real feedback, which [157] tackles by employing a survey. This work surveys different participants to assess the accuracy and significance of the refactorings produced by an explainable search-based refactoring (X-SBR) method for refactoring recommendations utilizing novel knowledge-informed MOEAs to direct the creation of the initial population, establish intelligent genetic operators, and elucidate the resulting refactoring solutions. Through the survey, the authors found that non-developers needed to be more knowledgeable, and their evaluation was not very useful.

Association rule mining (ARM), a type of ML method categorized under unsupervised learning, is used in [153] for metro map construction using the automatic information (information cartography) provided by the implication rules ($X \rightarrow Y$) from ARM databases together. Creating a metro map is an optimization problem that seeks the optimal metro lines inside the attribute graph, allowing for the visualization of feature and ARM information. The problem is addressed utilizing EAs, which apply to situations lacking unique insights and permit adaptive parameter configuration. The work pertains to the XAI sector, demonstrating that ARM information cartography effectively elucidates hidden knowledge inside ARM databases in an explainable manner via visualization.

Metaheuristic algorithms can also explain DL models, as [160] demonstrates by applying an evolutionary model for quality AR extraction to explain a long-short term memory (LSTM) DL model in a model-agnostic approach. Model-agnostic procedures are utilized to determine the outcomes or predictions derived from post-model training and independent of the model. This work intends to elucidate how LSTM produces time series forecasting values, referencing the input characteristics, i.e., historical data, enhancing DL models' efficacy and explainability for practitioners and stakeholders.

Explanation schemas are another approach to enhancing user understanding. In energy management systems, [149, 150] present recommendation systems that explain cause-effect relationships using multi-objective DE algorithms. Users can see the impact of their actions on energy systems and explore viable modifications to improve the energy system. in [148] a MOEA is proposed to optimize different explainability measures such as diversity, novelty, and interpretability to provide a recommendation system to end users. These recommendation lists are provided on the Pareto front given by the MOEA, showing the diversity and precision of these lists for each user.

The authors in [162] address the feedback problem verified in most analyzed works by incorporating the DM into an explainable evolutionary multi-objective optimization framework. They integrate learnable evolutionary models with interactive indicator-based optimization to create a learnable approach. ML models provide rule-based explanations to help the DM interpret the problem. The framework evolves dynamically, updating with DM feedback on solutions and rule-based descriptions, enhancing explainability. Similarly, [163] incorporates decision psychology into an interactive evolutionary framework. A dominance-based rough set approach generates "IF-THEN" rules reflecting DM preferences, guiding crossover mechanisms,

and optimizing Pareto front segments. From the perspective of behavioral psychology, these rules refine DM preferences, improving search processes.

These classifier systems are ideal for explaining to DMs and users how their actions and clauses affect the corresponding system when metaheuristic optimization algorithms are applied to obtain such information. Integrating the DM into these frameworks makes the system interactive, providing feedback from the DMs into the optimization processes and the provided rules, making the system more interpretable and more accurate with this expert knowledge integration.

7. Discussion and a use case in the energy field: the Explainergy Concept.

In the following, we go into the research questions formulated in Section 3.3, which are pivotal in our quest to generate a discussion topic. This topic revolves around integrating explainability approaches, as analyzed in this work, into metaheuristic optimization in the energy domain.

- **RQ1: Do XAI methods contribute to explaining metaheuristic performance behavior?** Indeed, XAI methods are often used to provide explanations for metaheuristic optimization. In the terms defined in XAI for explainability, methods such as SHAP and LIME for feature importance in ML can be applied to the context of metaheuristic optimization when it comes to understanding the most impactful landscape features in landscape analysis and even understanding hyperparameter tuning for algorithm configuration to interpret the influence of each metaheuristic parameter in its performance. Visualization, an extremely important XAI approach, is also integrated to observe the evolutionary processes and visually explain the observed search behavior. Most substantially, GP is an approach that intrinsically combines XAI concepts to provide intrinsic explanations in the evolutionary process since they can be visually perceived in a tree-like structure where their decision process can be interpreted, providing domain experts with insights into these algorithms.

- **RQ2: Do problem features influence metaheuristic diverse search procedures, and can their analysis have greater explainability?** Different problem features can influence metaheuristic performance, and also understanding the impact of problem variables

in metaheuristic behavior is extremely important since it gives insights into the problem, seeing what variables impact the algorithm's performance and providing essential information and what variables can be discarded. Looking at landscape features, they provide insights into the type of fitness and most influential features aiding metaheuristic design and how the algorithms explore the search space of the given problem. Moving into a feature space from the conventional parameter space is a novel approach that can illuminate the metaheuristic search process by promoting algorithm diversity, an important criterion in metaheuristic explainability.

- **RQ3: Can theoretical mathematical backgrounds provide understandable and transparent explanations for metaheuristic optimization?** Even though theoretical fundaments through mathematical formulations can provide substantial comprehension of the metaheuristic search procedures and their behaviors, they may not be sufficient to give non-experts in metaheuristic optimization, and the application domain since these explanations are extremely complex. These explanations interest metaheuristic experts with sufficient knowledge to understand the intrinsic processes of these algorithms, whereas metaheuristic experts need more detailed explanations. In contrast, non-experts need easy-to-understand and generally interpretable frameworks to understand the metaheuristic processes, which these mathematical fundaments usually need to provide.

- **RQ4: Can metaheuristics be utilized to introduce explanations in AI methods, namely ML models?** Regarding the use of metaheuristic algorithms to provide insights into ML models, it was shown that through the use of metaheuristics, black-box ML models can become more "white-box" models, where these algorithms are implemented in multiple circumstances, such as to extract counterfactuals and classifier systems via rule-based optimization, which are textual based explanations. GP is an intrinsic explainable EA that provides forecasting insights via individual construction. Knowledge graphs are also intrinsically and interpretable visually, where implementing metaheuristic algorithms can provide explicable recommendations to the DM in a specific domain.

Statistical analysis for state-of-the-art models shown in this review

presents a major challenge, where most works use an experimental analysis without any statistical significance when presenting their results. When dealing with metaheuristic optimizers, this application presents great relevance since these are stochastic algorithms with a random nature, where statistical validation instead of simulation-based validation presents better result confidence and, in addition, better explainability confidence as well. Still, works such as [36, 41, 46, 79, 91, 152] to cite a few, presented a statistical analysis, with methods such as the Kolmogorov–Smirnov test and t-test [172], the Taguchi method [173], the two-sample Anderson-Darling test [174] and the Kruskal–Wallis test [175], which very well established statistical analysis methods. By applying this validation, these works underpin the validity of their results and explanations.

In the analyzed works, a few implemented interaction mechanisms allowed expert feedback and explainability measures to be assessed regarding complexity, and the model was updated accordingly. This feedback is an important feature in that it allows expert knowledge in the metaheuristic framework through explainable features, mainly observed in rules and recommendations, which, through expert understanding, better interpretable solutions could be obtained.

Through different surveys and questionnaires to domain experts and non-experts, it was possible to conclude that experts require more details regarding the metaheuristic search processes and problem features. In contrast, non-experts require less detail in easier-to-understand explainable processes, which require less complex methods such as mathematical and theoretical explanations. Rule-based systems and visualizations can be better methods since they are simpler to understand. One is directly exposed in text via a rule structure, and the other in an explainable visual manner, which is user-understandable.

Even though real-world applications appear in the analyzed literature with some frequency, most of them do not appear in the author's field of expertise, which is the energy field. With the concepts acquired to try to provide explainable models to metaheuristic algorithms, the authors want to propose a discussion topic incorporating such methods and other possible explainable ventures to integrate these algorithms for optimization problems in the energy domain. The integration of such mechanisms would culminate in a concept named "Explainergy".

The concept of "Explainergy" emerges from integrating explainability into CI techniques. This pursuit of explainability on energy-related opti-

mization problems was first introduced in [114]. The concept builds on principles from XAI (see Fig. 9), adapting those to the domain of metaheuristics, which are often employed in energy optimization due to the complex characteristics inherent in such problems [1]. While explainability has been widely explored in areas like ML and AI [176], the Explainergy concept targets the interpretability of metaheuristic optimization methods, bridging the gap between advanced algorithmic techniques and their practical application in the energy sector for not only intrinsic metaheuristic understanding but also explain energy problem features influencing algorithm behavior [177, 178].

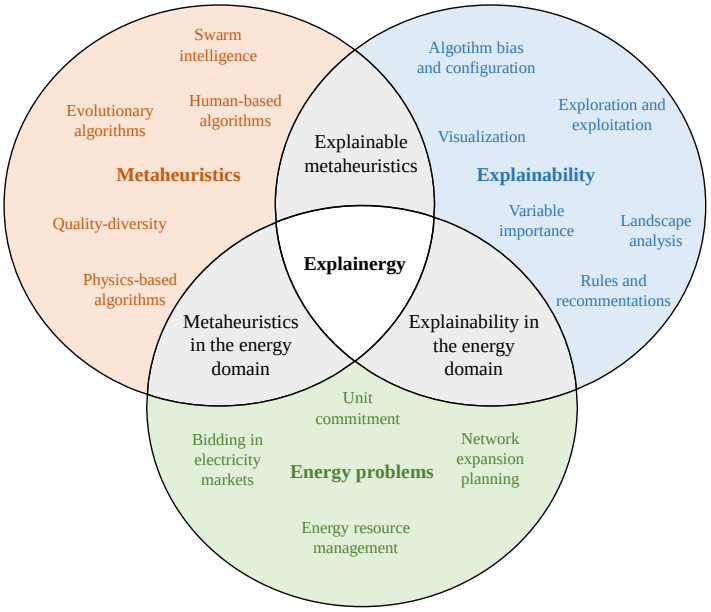

Figure 9: Concepts behind Explainergy idea.

Energy systems present numerous challenges that make optimization difficult. These include high-dimensional solution spaces, significant uncertainty (due to renewable energy generation, electric vehicle usage, and fluctuating loads), and time-sensitive decision-making. Metaheuristics are commonly used to address these challenges because they provide approximate solutions to complex problems efficiently [1]. However, their stochastic nature, variability in results across runs, and lack of theoretical guarantees on optimality can make metaheuristics less transparent and harder to trust for stakeholders unfamiliar with these approaches [114]. Explainergy aims to overcome these

limitations by incorporating explainability principles into the optimization process, making metaheuristics more interpretable, understandable, and, ultimately, more reliable for real-world applications.

One of the key advantages of incorporating explainability in metaheuristics lies in its ability to show how these algorithms explore and exploit the search space in a more understandable way to gain knowledge of the search mechanisms. Employing techniques for intrinsic explanations, such as algorithm bias and algorithm configuration (Sect. 4.1 and Sect. 4.2), visualization of search behavior (Sect. 4.5), to name a few methods and understanding problem characteristics such as variable importance and sensitivity analysis (Sect. 5.3), for example, users employing metaheuristics can gain insights into the DM processes. For example, transparent ML models (Sect. 6), such as DTs or surrogate models, can be integrated into the optimization framework to approximate how metaheuristics behave during the search process. These models can rank decision variables by importance, helping users understand which variables influence the outcome most and which can be discarded without affecting solution quality [179].

The Explainergy concept can also address the need to explain how parameter choices and problem structure impact the performance of metaheuristics. For instance, metaheuristic algorithms often incorporate starting solutions from deterministic techniques or hybridize operators from multiple algorithms. While effective, these design choices are often made with little discussion of their interpretability, leaving practitioners skeptical about how and why certain approaches work better for specific problems. Thus, embedding explainability into the algorithmic design process can provide a pathway for users to understand the working mechanism of metaheuristic algorithms, which is especially important in real-world energy applications such as optimal bidding in local markets [180], energy resource management [181], and real-time operational decisions [182].

Incorporating explainability in energy applications using metaheuristic optimization can be a key contribution to developing advanced optimization techniques. Evolutionary algorithms [183], estimation distribution algorithms [181], or swarm intelligence [184] have already demonstrated effectiveness in addressing the complexities present in energy optimization problems. By integrating explainability frameworks into these algorithms, the proposed "Explainergy" concept offers a new dimension of transparency, allowing users to achieve better solutions and understand the mechanisms and reasoning behind those solutions [185]. This is crucial to achieve trust and confidence in

adopting metaheuristic optimization methods for complex energy systems.

The work in [114] already started to tackle and implement some of these ideas into energy optimization, namely the optimal bidding in local electricity markets, which was previously analyzed in Section 5.3. Fig. 10 describes the framework proposed encompassing a post-optimization analysis for general mechanistic explanations by visualizing the solution encoding and the variable ranking, showing the variable influence in algorithm optimization. Still, this framework presents the first steps in applying some explainability to the energy domain, using this section's proposed "Explainergy" concepts.

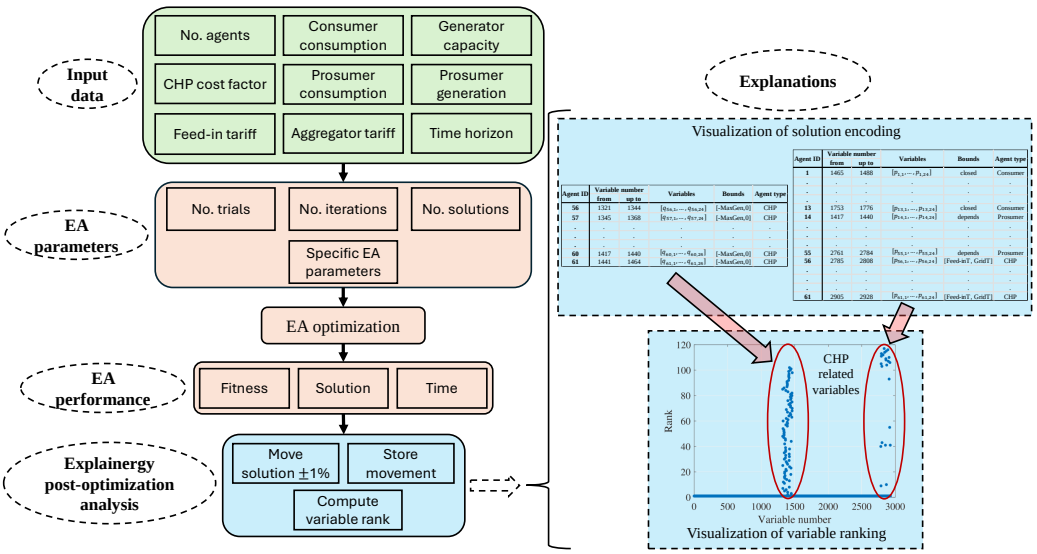

Figure 10: Framework encompassing explanations for metaheuristic optimization in the energy domain concerning variable ranking.

Taking inspiration from the analyzed topics and new ventures in explanations for metaheuristics in the energy domain, Fig. 11 shows an example of a rules and recommendations framework (Section 6.4) for an energy management problem, where the GP is used for optimization and knowledge from an aggregator, which is the DM, is considered as feedback from the optimization into the rule-making process. This framework can be a new approach for energy engineers and researchers in applying explanations to their models.

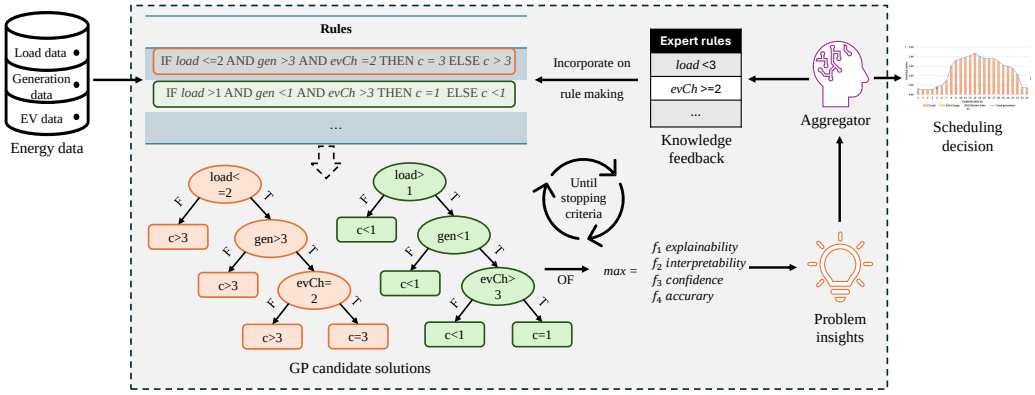

Figure 11: Example of rules and recommendation metaheuristic framework for an energy management problem.

8. Conclusions

This systematic study describes and explains the many approaches that seek to explain "black-box" metaheuristic optimization algorithms. The PRISMA technique is regarded in the study as the foundation of the extensive research strategy. Such interpretable techniques, their interaction with various issue landscapes, and their potential to increase the transparency of ML models were all made possible by this search approach. The results demonstrate the importance of XAI methods in offering insightful information about these optimizers' performance. Improved hyperparameter tweaking and deployment in practical situations are made possible by these techniques, which provide a deeper comprehension of algorithm behavior.

The paper also highlights how the inherent characteristics of optimization problems greatly impact metaheuristic algorithms' explainability and performance. Techniques for deciphering these interconnections and enhancing interpretability include landscape analysis and, more recently, QD. Furthermore, it is acknowledged that mathematical theory may provide clear explanations, and theoretical frameworks provide a strong basis for comprehending algorithms' behavior. However, there is still a significant gap between theoretical understanding and its practical application, which calls for more study to close. A particularly promising frontier identified in this review is the integration of metaheuristics to enhance the explainability of black-box ML models. By leveraging the optimization capabilities of metaheuristics, researchers can shed light on the inner workings of complex AI systems,

transforming them into more transparent and trustworthy tools. Moreover, the proposed "Explainergy" concept introduces an innovative framework for developing explainable metaheuristic algorithms tailored for large-scale optimization problems, particularly in the energy domain. This approach aims to enhance optimization models' transparency, usability, and practical relevance in addressing real-world challenges.

8.1. Research gaps and opportunities

The main research gap identified through the vast analyses of the literature regarding metaheuristic explainability is the small number of works considering domain knowledge and DM feedback in the developed models, which are generally independent of the application domain in which they are used. This integration can enhance the understanding and trust of these methods, which some works have already demonstrated through rule-based systems or knowledge graphs. Taking as an example, an energy engineer for an energy resource management problem resolution can give feedback and put prior knowledge regarding their expertise through existing models and solutions into the developed algorithm, which can help improve its mechanisms and explanation quality. Another research gap identified is the very little real-world application that these analyzed works presented, where most had simple benchmark functions and instances without testing the scalability of these methods. Real-world optimization and forecasting problems are generally large-scale with substantial amounts of data and uncertainty, needing the implementation of more complex CI optimization models, where explanations become crucial through the application of XAI with several local explanations to explain the various parts of the models, moving from simple global explanations, which may not perform well as the scalability and complexity of today's real-world problems in engineering, medical, financial and other domains.

The statistical analysis of results identifies a further research gap and a significant possibility for more investigation, especially in formulating precise benchmarking standards for metaheuristic optimization. This gap highlights the need for more rigorous evaluation frameworks that enhance the performance assessment of optimization algorithms and improve their explainability, guaranteeing more transparency and interpretability in the DM process.

A venue verified and can be taken advantage of in the literature uses multi-objective optimization to optimize and provide the users and developers with different explainable and understandable metrics, showing the

trade-off between the different objectives, which can be further explored with the proposal of new metrics, with these being model-specific and not model-agnostic. In this case, enhanced visualization approaches can be applied to provide a naive user with a simple to understand explanation with interactive capabilities, allowing users to navigate the complex search spaces of these techniques.

8.2. Future research

The review also identifies critical challenges and opportunities for future research, including:

- The need for regulatory agencies to standardize methods for evaluating explainability in metaheuristics;
- Bridging the gap between theoretical advancements and real-world applications, mainly in the energy domain;
- Expanding the use of explainable metaheuristics in diverse domains beyond energy systems, such as healthcare, finance, and transportation.

This systematic review wants to provide researchers, engineers, healthcare professionals, and others with tools to incorporate explainability into CI models, enabling the acceptance of these mechanisms, in which reluctance is still very high in their implementation due to their stochastic nature. For engineering problems in the energy domain, for example, metaheuristic methods are still badly viewed compared to mathematical optimization, which can guarantee a global optimum solution, and their implementation is very well grounded. Still, this review wants to encourage research in this area by providing the "Explainergy" framework so that research can provide new mechanisms for explanations, understanding, and trust of these mechanisms by practitioners in real-world implementations.

To sum up, this study establishes the foundation for developing explainability in metaheuristic optimization and emphasizes its importance for improving the usability and transparency of AI and optimization models. The conclusions and suggestions offer a path forward for scholars and professionals to fully utilize explainable metaheuristics, tackling urgent issues in practical problem-solving across various domains.

Acknowledgments

The work has been supported by Honda Research Institute Europe. In addition, the Portuguese team acknowledges the research facilities and conditions provided by GECAD, partially funded by Fundacao para a Ciencia e Tecnologia (FCT), namely grant UIDB/00760/2020 (https://doi.org/10.54499/UIDB/00760/2020). Joao Soares has an FCT grant, namely CEECIND/00420/2022 (doi:10.54499/2022.00420.CEECIND/CP1742/CT0001) to support his research work. José Almeida is supported by FCT with Ph.D. Grant 2022.09590.BD.

References

[1] João Soares, Tiago Pinto, Fernando Lezama, and Hugo Morais. Survey on Complex Optimization and Simulation for the New Power Systems Paradigm. *Complexity*, 2018:1–32, August 2018.

[2] Juan Guerra-Montenegro, Javier Sanchez-Medina, Ibai Laña, David Sanchez-Rodriguez, Itziar Alonso-Gonzalez, and Javier Del Ser. Computational Intelligence in the hospitality industry: A systematic literature review and a prospect of challenges. *Applied Soft Computing*, 102:107082, April 2021.

[3] Kumeshan Reddy and Akshay K. Saha. A review of swarm-based metaheuristic optimization techniques and their application to doubly fed induction generator. *Heliyon*, 8(10):e10956, 2022.

[4] Zhongqiang Ma, Guohua Wu, Ponnuthurai Nagaratnam Suganthan, Aijuan Song, and Qizhang Luo. Performance assessment and exhaustive listing of 500+ nature-inspired metaheuristic algorithms. *Swarm and Evolutionary Computation*, 77, 3 2023.

[5] Armin Nabaei, Melika Hamian, Mohammad Reza Parsaei, Reza Safdari, Taha Samad-Soltani, Houman Zarrabi, and A. Ghassemi. Topologies and performance of intelligent algorithms: a comprehensive review. *Artificial Intelligence Review*, 49:79–103, 1 2018.

[6] Eneko Osaba, Esther Villar-Rodriguez, Javier Del Ser, Antonio J. Nebro, Daniel Molina, Antonio LaTorre, Ponnuthurai N. Suganthan, Carlos A. Coello Coello, and Francisco Herrera. A tutorial on the design,

experimentation and application of metaheuristic algorithms to real-world optimization problems. *Swarm and Evolutionary Computation*, 64:100888, 2021.

[7] Kenneth Sörensen. Metaheuristics-the metaphor exposed. *International Transactions in Operational Research*, 22:3–18, 1 2015.

[8] Zbigniew Michalewicz. Ubiquity symposium: Evolutionary computation and the processes of life: the emperor is naked: evolutionary algorithms for real-world applications. *Ubiquity*, 2012, November 2012.

[9] James M. Whitacre. Survival of the flexible: Explaining the recent popularity of nature-inspired optimization within a rapidly evolving world. *Computing*, 93:135–146, 12 2011.

[10] Kalyanmoy Deb and Matthias Ehrgott. On generalized dominance structures for multi-objective optimization. *Mathematical and Computational Applications*, 28:100, 10 2023.

[11] Trung Thanh Nguyen, Shengxiang Yang, and Juergen Branke. Evolutionary dynamic optimization: A survey of the state of the art. *Swarm and Evolutionary Computation*, 6:1–24, 10 2012.

[12] Gesina Schwalbe and Bettina Finzel. A comprehensive taxonomy for explainable artificial intelligence: a systematic survey of surveys on methods and concepts. *Data Mining and Knowledge Discovery*, 2023.

[13] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 6 2020.

[14] Katherine Mary Malan. A survey of advances in landscape analysis for optimisation. *Algorithms*, 14, 2 2021.

[15] Changwu Huang, Yuanxiang Li, and Xin Yao. A survey of automatic parameter tuning methods for metaheuristics. *IEEE Transactions on Evolutionary Computation*, 24:201–216, 4 2020.

[16] Olusanmi Hundogan, Xixi Lu, Yupei Du, and Hajo A. Reijers. Created: Generating viable counterfactual sequences for predictive process analytics. In Marta Indulska, Iris Reinhartz-Berger, Carlos Cetina, and Oscar Pastor, editors, *Advanced Information Systems Engineering*, pages 541–557, Cham, 2023. Springer Nature Switzerland.

[17] Shingo Mabu, Masanao Obayashi, and Takashi Kuremoto. Ensemble learning of rule-based evolutionary algorithm using multi layer perceptron for stock trading models. In *2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS)*, pages 624–629, 2014.

[18] Yi Mei, Qi Chen, Andrew Lensen, Bing Xue, and Mengjie Zhang. Explainable artificial intelligence by genetic programming: A survey. *IEEE Transactions on Evolutionary Computation*, 27:621–641, 6 2023.

[19] Daniel Howard and Mark A. Edwards. Explainable a.i.: The promise of genetic programming multi-run subtree encapsulation. In *2018 International Conference on Machine Learning and Data Engineering (iCMLDE)*, pages 158–159, 2018.

[20] Jaume Bacardit, Alexander E. I. Brownlee, Stefano Cagnoni, Giovanni Iacca, John McCall, and David Walker. The intersection of evolutionary computation and explainable ai. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, GECCO '22, page 1757–1762, New York, NY, USA, 2022. Association for Computing Machinery.

[21] Ryan Zhou, Jaume Bacardit, Alexander E. I. Brownlee, Stefano Cagnoni, Martin Fyvie, Giovanni Iacca, John McCall, Niki van Stein, David J. Walker, and Ting Hu. Evolutionary computation and explainable ai: A roadmap to understandable intelligent systems. *IEEE Transactions on Evolutionary Computation*, pages 1–1, 2024.

[22] Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M.

Lalu, Tianjing Li, Elizabeth W. Loder, Evan Mayo-Wilson, Steve McDonald, Luke A. McGuinness, Lesley A. Stewart, James Thomas, Andrea C. Tricco, Vivian A. Welch, Penny Whiting, and David Moher. The prisma 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ*, 372, 3 2021.

[23] Seyedali Mirjalili, Seyed Mohammad Mirjalili, and Andrew Lewis. Grey wolf optimizer. *Advances in Engineering Software*, 69:46–61, 2014.

[24] Xin-She Yang and Suash Deb. Cuckoo search via lévy flights. In *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, pages 210–214, 2009.

[25] Seth Bullock. Smooth operator? understanding and visualising mutation bias. In Jozef Kelemen and Petr Sosík, editors, *Advances in Artificial Life*, pages 602–612, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.

[26] Liviu Panait, R. Paul Wiegand, and Sean Luke. A sensitivity analysis of a cooperative coevolutionary algorithm biased for optimization. In Kalyanmoy Deb, editor, *Genetic and Evolutionary Computation – GECCO 2004*, pages 573–584, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[27] Sang Moon Soak, David W. Corne, and Byung Ha Ahn. The edge-window-decoder representation for tree-based problems. *IEEE Transactions on Evolutionary Computation*, 10:124–144, 4 2006.

[28] Ting Hu, Marco Tomassini, and Wolfgang Banzhaf. A network perspective on genotype–phenotype mapping in genetic programming. *Genetic Programming and Evolvable Machines*, 21:375–397, 9 2020.

[29] Diederick Vermetten, Anna V. Kononova, Fabio Caraffini, Hao Wang, and Thomas Bäck. Is there anisotropy in structural bias? In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, GECCO '21, page 1243–1250, New York, NY, USA, 2021. Association for Computing Machinery.

[30] Diederick Vermetten, Fabio Caraffini, Bas van Stein, and Anna V. Kononova. Using structural bias to analyse the behaviour of modular

cma-es. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, GECCO '22, page 1674–1682, New York, NY, USA, 2022. Association for Computing Machinery.

[31] Bas Van Stein, Diederick Vermetten, Fabio Caraffini, and Anna V. Kononova. Deep bias: Detecting structural bias using explainable ai. In *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, GECCO '23 Companion, page 455–458, New York, NY, USA, 2023. Association for Computing Machinery.

[32] Niki van Stein, Diederick Vermetten, Anna V. Kononova, and Thomas Bäck. Explainable benchmarking for iterative optimization heuristics. 1 2024.

[33] James Montgomery. Crossover and the different faces of differential evolution searches. In *IEEE Congress on Evolutionary Computation*, pages 1–8, 2010.

[34] S. K. Smit and A. E. Eiben. Parameter tuning of evolutionary algorithms: Generalist vs. specialist. In *Applications of Evolutionary Computation*, volume 6024 LNCS, pages 542–551. Springer Berlin Heidelberg, 2010.

[35] A. Ortiz, J. A. Cabrera, F. Nadal, and A. Bonilla. Dimensional synthesis of mechanisms using differential evolution with auto-adaptive control parameters. *Mechanism and Machine Theory*, 64:210–229, 2013.

[36] Krzysztof Michalak. The effects of asymmetric neighborhood assignment in the moea/d algorithm. *Applied Soft Computing Journal*, 25:97–106, 2014.

[37] Shane Strasser and John W. Sheppard. Evaluating factored evolutionary algorithm performance on binary deceptive functions. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–8, 2017.

[38] Shu-Yu Kuo, Ching-Hsuan Wu, Cheng-Chun Chen, and Yao-Hsin Chou. A novel metaheuristic: Fast jaguar algorithm. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 146–151, 2021.

[39] Ran Wang and Zichao Zhang. Set theory-based operator design in evolutionary algorithms for solving knapsack problems. *IEEE Transactions on Evolutionary Computation*, 25:1133–1147, 12 2021.

[40] Ching-Hsuan Wu, Jyun-Yi Shen, Pei-Shin Huang, Cheng-Yen Hua, Yu-Chi Jiang, Shu-Yu Kuo, and Yao-Hsin Chou. A novel explainable nature-inspired metaheuristic: Jaguar algorithm with precision hunting behavior. In *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1494–1499, 2022.

[41] Anabel Martinez-Vargas, Jesus A. Gomez-Aviles, M. A. Cosio-Leon, and Angel G. Andrade. Explaining the walking through of a team of algorithms. *Computer*, 56:67–81, 2 2023.

[42] Steven Halim and Hoong Chuin Lau. *Tuning Tabu Search Strategies Via Visual Diagnosis*, pages 365–388. Springer US, Boston, MA, 2007.

[43] Anas Al-Tirawi and Robert G. Reynolds. How to design a trustable cultural algorithm using common value auctions. In *2021 Third International Conference on Transdisciplinary AI (TransAI)*, pages 74–84, 2021.

[44] Anas Al-Tirawi and Robert G. Reynolds. Cultural algorithms as a framework for the design of trustable evolutionary algorithms. *International Journal of Semantic Computing*, 16(01):107–134, 2022.

[45] Marcelo de Souza and Marcus Ritt. Improved regression models for algorithm configuration. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '22, page 222–231, New York, NY, USA, 2022. Association for Computing Machinery.

[46] Muath Ibrahim Jarrah, A. S.M. Jaya, Zakaria N. Alqattan, Mohd Asyadi Azam, Rosni Abdullah, Hazim Jarrah, and Ahmed Ismail Abu-Khadrah. A novel explanatory hybrid artificial bee colony algorithm for numerical function optimization. *Journal of Supercomputing*, 76:9330–9354, 12 2020.

[47] Michal Pluhacek, Anezka Kazikova, Tomas Kadavy, Adam Viktorin, and Roman Senkerik. Explaining soma: the relation of stochastic perturbation to population diversity and parameter space coverage. In

Proceedings of the Genetic and Evolutionary Computation Conference Companion, GECCO '21, page 1944–1952, New York, NY, USA, 2021. Association for Computing Machinery.

[48] R.C. Purshouse and P.J. Fleming. Evolutionary many-objective optimisation: an exploratory analysis. In *The 2003 Congress on Evolutionary Computation, 2003. CEC '03.*, volume 3, pages 2066–2073 Vol.3, 2003.

[49] Robin C. Purshouse and Peter J. Fleming. On the evolutionary optimization of many conflicting objectives. *IEEE Transactions on Evolutionary Computation*, 11:770–784, 12 2007.

[50] Stephen Chen, Antonio Bolufé-Röhler, James Montgomery, and Tim Hendtlass. An analysis on the effect of selection on exploration in particle swarm optimization and differential evolution. In *2019 IEEE Congress on Evolutionary Computation (CEC)*, pages 3037–3044, 2019.

[51] Alexandre Bettinger, Armelle Brun, and Anne Boyer. Independent influence of exploration and exploitation for metaheuristic-based recommendations. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, GECCO '22, page 475–478, New York, NY, USA, 2022. Association for Computing Machinery.

[52] Alexandre Bettinger, Armelle Brun, and Anne Boyer. Presence variations for e&e transitions explainability. *Procedia Computer Science*, 207:1802–1811, 2022.

[53] Kashif Hussain, Mohd Najib Mohd Salleh, Shi Cheng, and Yuhui Shi. On the exploration and exploitation in popular swarm-based metaheuristic algorithms. *Neural Computing and Applications*, 31:7665–7683, 11 2019.

[54] Amreek Singh and Kusum Deep. Exploration–exploitation balance in artificial bee colony algorithm: a critical analysis. *Soft Computing*, 23:9525–9536, 10 2019.

[55] J. A. Vázquez-Rodríguez and S. Petrovic. A mixture experiments multi-objective hyper-heuristic. *Journal of the Operational Research Society*, 64:1664–1675, 2013.

[56] Fei Liu, Xialiang Tong, Mingxuan Yuan, Xi Lin, Fu Luo, Zhenkun Wang, Zhichao Lu, and Qingfu Zhang. Evolution of heuristics: towards efficient automatic algorithm design using large language model. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

[57] Haoran Ye, Jiarui Wang, Zhiguang Cao, Federico Berto, Chuanbo Hua, HAEYEON KIM, Jinkyoo Park, and Guojie Song. Reevo: Large language models as hyper-heuristics with reflective evolution. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 43571–43608. Curran Associates, Inc., 2024.

[58] Y T Sasmi Hidayatul, Arif Djunaidy, and Ahmad Muklason. Solving multi-objective vehicle routing problem using hyper-heuristic method by considering balance of route distances. In *2019 International Conference on Information and Communications Technology (ICOIACT)*, pages 937–942, 2019.

[59] Melissa Sanchez, Jorge M. Cruz-Duarte, Jose C. Ortiz-Bayliss, and Ivan Amaya. Sequence-based selection hyper-heuristic model via map-elites. *IEEE Access*, 9:116500–116527, 2021.

[60] Helio J. C. Barbosa, Heder S. Bernardino, and André M.S. Barreto. Using performance profiles to analyze the results of the 2006 cec constrained optimization competition. In *IEEE Congress on Evolutionary Computation*, pages 1–8, 2010.

[61] Eric Medvet, Marco Virgolin, Mauro Castelli, Peter A.N. Bosman, Ivo Gonçalves, and Tea Tušar. Unveiling evolutionary algorithm representation with du maps. *Genetic Programming and Evolvable Machines*, 19:351–389, 9 2018.

[62] Mathew J. Walter, David J. Walker, and Matthew J. Craven. An explainable visualisation of the evolutionary search process. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, GECCO '22, page 1794–1802, New York, NY, USA, 2022. Association for Computing Machinery.

[63] Mathew J. Walter, David J. Walker, and Matthew J. Craven. Visualizing population dynamics to examine algorithm performance. *IEEE Transactions on Evolutionary Computation*, 26:1501–1510, 12 2022.

[64] Tomoaki Takagi, Keiki Takadama, and Hiroyuki Sato. Directional pareto front and its estimation to encourage multi-objective decision-making. *IEEE Access*, 11:20619–20634, 2023.

[65] Shu Yu Kuo, Yu Chi Jiang, Cheng Yen Hua, Yao Hsin Chou, and Sy Yen Kuo. Evo-panel: Dynamic visualization tool for optimization process. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 7:1717–1732, 12 2023.

[66] Yansong Huang, Zherui Zhang, Ao Jiao, Yuxin Ma, and Ran Cheng. A comparative visual analytics framework for evaluating evolutionary processes in multi-objective optimization. *IEEE Transactions on Visualization and Computer Graphics*, 30:661–671, 1 2024.

[67] Ilia Stepin, Jose M. Alonso-Moral, Alejandro Catala, and Martín Pereira-Fariña. An empirical study on how humans appreciate automated counterfactual explanations which embrace imprecise information. *Information Sciences*, 618:379–399, 12 2022.

[68] John H. Drake, Ahmed Kheiri, Ender Özcan, and Edmund K. Burke. Recent advances in selection hyper-heuristics. *European Journal of Operational Research*, 285:405–428, 9 2020.

[69] Angus Kenny, Tapabrata Ray, Steffen Limmer, Hemant Kumar Singh, Tobias Rodemann, and Markus Olhofer. A hierarchical dissimilarity metric for automated machine learning pipelines, and visualizing search behaviour. In *Applications of Evolutionary Computation: 27th European Conference, EvoApplications 2024, Held as Part of EvoStar 2024, Aberystwyth, UK, April 3–5, 2024, Proceedings, Part II*, page 115–129, Berlin, Heidelberg, 2024. Springer-Verlag.

[70] David E Golberg. Genetic algorithms in search, optimization, and machine learning. *Addion wesley*, 1989(102):36, 1989.

[71] Hans-Georg Beyer. An alternative explanation for the manner in which genetic algorithms operate. *Biosystems*, 41:1–15, 1 1997.

[72] Yang Yu and Zhi Hua Zhou. A new approach to estimating the expected first hitting time of evolutionary algorithms. *Artificial Intelligence*, 172:1809–1832, 10 2008.

[73] Alexander Raß, Manuel Schmitt, and Rolf Wanka. Explanation of stagnation at points that are not local optima in particle swarm optimization by potential analysis. In *Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation*, GECCO Companion '15, page 1463–1464, New York, NY, USA, 2015. Association for Computing Machinery.

[74] Duc Cuong Dang, Thomas Jansen, and Per Kristian Lehre. Populations can be essential in tracking dynamic optima. *Algorithmica*, 78:660–680, 6 2017.

[75] Simon Jones, Alan F. Winfield, Sabine Hauert, and Matthew Studley. Onboard evolution of understandable swarm behaviors. *Advanced Intelligent Systems*, 1, 10 2019.

[76] Junaid Akhtar, Mian M. Awais, and Basit B. Koshul. Evolutionary algorithms based on non-darwinian theories of evolution. In *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, pages 2554–2560, 2008.

[77] Vitor Nazário Coelho, Igor Machado Coelho, Nenad Mladenović, Helena Ramalhinho, Luiz Satoru Ochi, Frederico G. Guimarães, and Marcone J.F. Souza. Less is more: The neighborhood guided evolution strategies convergence on some classic neighborhood operators. In *Variable Neighborhood Search*, volume 11328 LNCS, pages 77–88. Springer International Publishing, 2019.

[78] Ricardo H.C. Takahashi. On the convergence of decomposition algorithms in many-objective problems. In *Evolutionary Multi-Criterion Optimization*, volume 11411 LNCS, pages 39–50. Springer International Publishing, 2019.

[79] Tome Eftimov, Gorjan Popovski, Dragi Kocev, and Peter Korošec. Performance2vec: a step further in explainable stochastic optimization algorithm performance. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*, GECCO '20, page

193–194, New York, NY, USA, 2020. Association for Computing Machinery.

[80] Ruan G.S. Gomes, Guilherme J.C. Gomes, and Jasper A. Vrugt. A hybrid multi-step sensitivity-driven evolutionary polynomial regression enables robust model structure selection. *Engineering Applications of Artificial Intelligence*, 11 2022.

[81] Zhaobo Liu, Guo Li, Haili Zhang, Zhengping Liang, and Zexuan Zhu. Multifactorial evolutionary algorithm based on diffusion gradient descent. *IEEE Transactions on Cybernetics*, 2023.

[82] Chao Qian, Yang Yu, Ke Tang, Xin Yao, and Zhi Hua Zhou. Maximizing submodular or monotone approximately submodular functions by multi-objective evolutionary algorithms. *Artificial Intelligence*, 275:279–294, 10 2019.

[83] Chao Qian, Dan Xuan Liu, Chao Feng, and Ke Tang. Multi-objective evolutionary algorithms are generally good: Maximizing monotone submodular functions over sequences. *Theoretical Computer Science*, 943:241–266, 1 2023.

[84] Y. Katada and K. Ohkura. Estimating the degree of neutrality in fitness landscapes by the nei's standard genetic distance - an application to evolutionary robotics. In *2006 IEEE International Conference on Evolutionary Computation*, pages 483–490, 2006.

[85] Marek Kubiak. Distance measures and fitness-distance analysis for the capacitated vehicle routing problem. *Operations Research/ Computer Science Interfaces Series*, 39:345–364, 2007.

[86] Silja Meyer-Nieberg and Hans-Georg Beyer. Why noise may be good: additive noise on the sharp ridge. In *Proceedings of the 10th annual conference on Genetic and evolutionary computation*, pages 511–518. ACM, 7 2008.

[87] Sang Moon Soak and Moongu Jeon. The property analysis of evolutionary algorithms applied to spanning tree problems. *Applied Intelligence*, 32:96–121, 2 2010.

[88] Alhandra C. M. A. Tepedino, Ricardo H. C. Takahashi, and Eduardo Gontijo Carrano. Distance based nsga-ii for earliness and tardiness minimization in parallel machine scheduling. In *2013 IEEE Congress on Evolutionary Computation*, pages 317–324, 2013.

[89] Marie Éléonore Marmion, Laetitia Jourdan, and Clarisse Dhaenens. Fitness landscape analysis and metaheuristics efficiency. *Journal of Mathematical Modelling and Algorithms*, 12:3–26, 2013.

[90] Devin Greene and Kristina Crona. The changing geometry of a fitness landscape along an adaptive walk. *PLoS Computational Biology*, 10, 2014.

[91] Mang Wang, Bin Li, Guofu Zhang, and Xin Yao. Population evolvability: Dynamic fitness landscape analysis for population-based metaheuristic algorithms. *IEEE Transactions on Evolutionary Computation*, 22:550–563, 8 2018.

[92] Ana Nikolikj, Risto Trajanov, Gjorgjina Cenikj, Peter Korošec, and Tome Eftimov. Identifying minimal set of exploratory landscape analysis features for reliable algorithm performance prediction. In *2022 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–8, 2022.

[93] Ana Nikolikj, Ryan Lang, Peter Korošec, and Tome Eftimov. Explaining differential evolution performance through problem landscape characteristics. In *Bioinspired Optimization Methods and Their Applications*, volume 13627 LNCS, pages 99–113. Springer International Publishing, 2022.

[94] Ana Kostovska, Diederick Vermetten, Sašo Džeroski, Carola Doerr, Peter Korosec, and Tome Eftimov. The importance of landscape features for performance prediction of modular cma-es variants. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '22, page 648–656, New York, NY, USA, 2022. Association for Computing Machinery.

[95] Chloe M. Barnes, Abida Ghouri, and Peter R. Lewis. Explaining evolutionary agent-based models via principled simplification. *Artificial Life*, 27:143–163, 3 2022.

[96] Sebastian Herrmann, Gabriela Ochoa, and Franz Rothlauf. Coarse-grained barrier trees of fitness landscapes. In Julia Handl, Emma Hart, Peter R. Lewis, Manuel López-Ibáñez, Gabriela Ochoa, and Ben Paechter, editors, *Parallel Problem Solving from Nature – PPSN XIV*, pages 901–910, Cham, 2016. Springer International Publishing.

[97] Sarah L. Thomson, Gabriela Ochoa, Sébastien Verel, and Nadarajen Veerapen. Inferring future landscapes: Sampling the local optima level. *Evolutionary Computation*, 28:621–641, 2019.

[98] Sarah L. Thomson and Gabriela Ochoa. On funnel depths and acceptance criteria in stochastic local search. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '22, page 287–295, New York, NY, USA, 2022. Association for Computing Machinery.

[99] Jean-Baptiste Mouret and Jeff Clune. Illuminating search spaces by mapping elites. 4 2015.

[100] Joshua E. Auerbach, Giovanni Iacca, and Dario Floreano. Gaining insight into quality diversity. In *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion*, GECCO '16 Companion, page 1061–1064, New York, NY, USA, 2016. Association for Computing Machinery.

[101] Justin K. Pugh, Lisa B. Soros, and Kenneth O. Stanley. Quality diversity: A new frontier for evolutionary computation. *Frontiers Robotics AI*, 3, 7 2016.

[102] Vassilis Vassiliades, Konstantinos Chatzilygeroudis, and Jean Baptiste Mouret. Using centroidal voronoi tessellations to scale up the multidimensional archive of phenotypic elites algorithm. *IEEE Transactions on Evolutionary Computation*, 22:623–630, 8 2018.

[103] Neil Urquhart, Michael Guckert, and Simon Powers. Increasing trust in meta-heuristics by using map-elites. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, GECCO '19, page 1345–1348, New York, NY, USA, 2019. Association for Computing Machinery.

[104] Neil Urquhart, Silke Höhl, and Emma Hart. Automated, explainable rule extraction from map-elites archives. In *Applications of Evolutionary Computation*, volume 12694 LNCS, pages 258–272. Springer International Publishing, 2021.

[105] Ludovico Scarton and Alexander Hagg. On the suitability of representations for quality diversity optimization of shapes. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '23, page 963–971, New York, NY, USA, 2023. Association for Computing Machinery.

[106] Kalyan Veeramachaneni, Ekaterina Vladislavleva, and Una May O'Reilly. Knowledge mining sensory evaluation data: Genetic programming, statistical techniques, and swarm optimization. *Genetic Programming and Evolvable Machines*, 13:103–133, 3 2012.

[107] Valentina Richthammer, Tobias Scheinert, and Michael Glaß. Data mining in system-level design space exploration of embedded systems. In *Embedded Computer Systems: Architectures, Modeling, and Simulation*, volume 12471 LNCS, pages 52–66. Springer International Publishing, 2020.

[108] Martin Fyvie, John A.W. McCall, and Lee A. Christie. Towards explainable metaheuristics: Pca for trajectory mining in evolutionary algorithms. In *Artificial Intelligence XXXVIII*, volume 13101 LNAI, pages 89–102. Springer Science and Business Media Deutschland GmbH, 2021.

[109] Aidan Wallace, Alexander E.I. Brownlee, and David Cairns. Towards explaining metaheuristic solution quality by data mining surrogate fitness models for importance of variables. In *Artificial Intelligence XXXVIII*, volume 13101 LNAI, pages 58–72. Springer Science and Business Media Deutschland GmbH, 2021.

[110] Alexander E I Brownlee, Aidan Wallace, and David Cairns. Mining markov network surrogates to explain the results of metaheuristic optimisation. In *Proceedings of the SICSA eXplainable Artifical Intelligence Workshop*, pages 1–7, 2021.

[111] Manjinder Singh, Alexander E. I. Brownlee, and David Cairns. Towards explainable metaheuristic: mining surrogate fitness models for importance of variables. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, GECCO '22, page 1785–1793, New York, NY, USA, 2022. Association for Computing Machinery.

[112] Martin Fyvie, John A.W. McCall, Lee A. Christie, Alexander E.I. Brownlee, and Manjinder Singh. Towards explainable metaheuristics: Feature extraction from trajectory mining. *Expert Systems*, 2023.

[113] Martin Fyvie, John Mccall, Lee Christie, and Alexander Brownlee. Explaining a staff rostering genetic algorithm using sensitivity analysis and trajectory analysis. In *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, GECCO '23 Companion, page 1648–1656, New York, NY, USA, 2023. Association for Computing Machinery.

[114] Fernando Lezama, Jose Almeida, Joao Soares, and Zita Vale. Explainergy: Towards explainability of metaheuristic performance in the energy field. In *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 783–788, 2023.

[115] Rahul Venkatesh, Eric Wong, and Zico Kolter. Adversarial robustness in discontinuous spaces via alternating sampling & descent. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 4651–4660, 2023.

[116] Sarah L. Thomson, Jason Adair, Alexander E. I. Brownlee, and Daan van den Berg. From fitness landscapes to explainable ai and back. In *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, GECCO '23 Companion, page 1663–1667, New York, NY, USA, 2023. Association for Computing Machinery.

[117] Shinichi Shirakawa and Tomoharu Nagao. Bag of local landscape features for fitness landscape analysis. *Soft Computing*, 20(10):3787–3802, October 2016.

[118] Hendrik Richter. Codynamic fitness landscapes of coevolutionary minimal substrates. In *2014 IEEE Congress on Evolutionary Computation (CEC)*, pages 2692–2699, Beijing, China, July 2014. IEEE.

[119] Katherine M. Malan, Johannes F. Oberholzer, and Andries P. Engelbrecht. Characterising constrained continuous optimisation problems. In *2015 IEEE Congress on Evolutionary Computation (CEC)*, pages 1351–1358, Sendai, Japan, May 2015. IEEE.

[120] Fabio Caraffini, Ferrante Neri, and Lorenzo Picinali. An analysis on separability for Memetic Computing automatic design. *Information Sciences*, 265:1–22, May 2014.

[121] Olaf Mersmann, Bernd Bischl, Heike Trautmann, Mike Preuss, Claus Weihs, and Günter Rudolph. Exploratory landscape analysis. In *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, GECCO '11, page 829–836, New York, NY, USA, 2011. Association for Computing Machinery.

[122] Rachael Morgan and Marcus Gallagher. Length Scale for Characterising Continuous Optimization Problems. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Carlos A. Coello Coello, Vincenzo Cutello, Kalyanmoy Deb, Stephanie Forrest, Giuseppe Nicosia, and Mario Pavone, editors, *Parallel Problem Solving from Nature - PPSN XII*, volume 7491, pages 407–416. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. Series Title: Lecture Notes in Computer Science.

[123] Arnaud Liefooghe, Fabio Daolio, Sebastien Verel, Bilel Derbel, Hernan Aguirre, and Kiyoshi Tanaka. Landscape-Aware Performance Prediction for Evolutionary Multiobjective Optimization. *IEEE Transactions on Evolutionary Computation*, 24(6):1063–1077, December 2020.

[124] Gabriela Ochoa, Marco Tomassini, Sebástien Vérel, and Christian Darabos. A study of NK landscapes' basins and local optima networks. In *Proceedings of the 10th annual conference on Genetic and evolutionary computation*, pages 555–562, Atlanta GA USA, July 2008. ACM.

[125] Anna Sergeevna Bosman, Andries Engelbrecht, and Mardé Helbig. Visualising basins of attraction for the cross-entropy and the squared error

neural network loss functions. *Neurocomputing*, 400:113–136, August 2020.

[126] Yuan Sun, Michael Kirley, and Saman K. Halgamuge. Quantifying Variable Interactions in Continuous Optimization Problems. *IEEE Transactions on Evolutionary Computation*, 21(2):249–264, April 2017.

[127] Quentin Renau, Johann Dreo, Carola Doerr, and Benjamin Doerr. Towards explainable exploratory landscape analysis: Extreme feature selection for classifying bbob functions. In Pedro A. Castillo and Juan Luis Jiménez Laredo, editors, *Applications of Evolutionary Computation*, pages 17–33, Cham, 2021. Springer International Publishing.

[128] Raphaël Cosson, Bilel Derbel, Arnaud Liefooghe, Sébastien Verel, Hernan Aguirre, Qingfu Zhang, and Kiyoshi Tanaka. Cost-vs-accuracy of sampling in multi-objective combinatorial exploratory landscape analysis. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '22, page 493–501, New York, NY, USA, 2022. Association for Computing Machinery.

[129] Deepak Venugopal, Max Garzon, Nirman Kumar, Ching-Chi Yang, and Lih-Yuan Deng. *Metaheuristics of DR Methods*, pages 199–218. Springer International Publishing, Cham, 2022.

[130] Alejandro Barredo-Arrieta and Javier Del Ser. Plausible counterfactuals: Auditing deep learning classifiers with realistic adversarial examples. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7, 2020.

[131] Jacqueline Höllig, Cedric Kulbach, and Steffen Thoma. Tsevo: Evolutionary counterfactual explanations for time series classification. In *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 29–36, 2022.

[132] Raphael Mazzine Barbosa de Oliveira, Kenneth Sörensen, and David Martens. A model-agnostic and data-independent tabu search algorithm to generate counterfactuals for tabular, image, and text data. *European Journal of Operational Research*, 2023.

[133] Hayden Andersen, Andrew Lensen, Will Browne, and Yi Mei. Producing diverse rashomon sets of counterfactual explanations with niching

particle swarm optimization algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '23, page 393–401, New York, NY, USA, 2023. Association for Computing Machinery.

[134] Hongqing Cao, Friedrich Recknagel, Gea Jae Joo, and Dong Kyun Kim. Discovery of predictive rule sets for chlorophyll-a dynamics in the nakdong river (korea) by means of the hybrid evolutionary algorithm hea. *Ecological Informatics*, 1:43–53, 1 2006.

[135] Wai Sum Chan, Friedrich Recknagel, Hongqing Cao, and Ho Dong Park. Elucidation and short-term forecasting of microcystin concentrations in lake suwa (japan) by means of artificial neural networks and evolutionary algorithms. *Water Research*, 41:2247–2255, 2007.

[136] Hongqing Cao, Friedrich Recknagel, and Philip T. Orr. Parameter optimization algorithms for evolving rule models applied to freshwater ecosystems. *IEEE Transactions on Evolutionary Computation*, 18:793–806, 12 2014.

[137] Jayashree Chadalawada, Vojtech Havlicek, and Vladan Babovic. A genetic programming approach to system identification of rainfall-runoff models. *Water Resources Management*, 31:3975–3992, 9 2017.

[138] Rosa Senatore, Antonio Della Cioppa, and Angelo Marcelli. Automatic diagnosis of parkinson disease through handwriting analysis: A cartesian genetic programming approach. In *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, pages 312–317, 2019.

[139] Lijie Xie, Zhaoming Hu, Xingjuan Cai, Wensheng Zhang, and Jinjun Chen. Explainable recommendation based on knowledge graph and multi-objective optimization. *Complex and Intelligent Systems*, 7:1241–1252, 6 2021.

[140] Xingjuan Cai, Lijie Xie, Rui Tian, and Zhihua Cui. Explicable recommendation based on knowledge graph. *Expert Systems with Applications*, 200, 8 2022.

[141] Stefan Heindorf, Lukas Blübaum, Nick Düsterhus, Till Werner, Varun Nandkumar Golani, Caglar Demir, and Axel-Cyrille

Ngonga Ngomo. Evolearner: Learning description logics with evolutionary algorithms. In *Proceedings of the ACM Web Conference 2022*, WWW '22, page 818–828, New York, NY, USA, 2022. Association for Computing Machinery.

[142] Zachariah Phillips and Kelly Cohen. Genetically trained fuzzy cognitive maps for effects based operations. In Scott Dick, Vladik Kreinovich, and Pawan Lingras, editors, *Applications of Fuzzy Techniques*, pages 185–195, Cham, 2023. Springer International Publishing.

[143] R. E. Smith and Max Kun Jiang. Milcs: a mutual information learning classifier system. In *Proceedings of the 9th Annual Conference Companion on Genetic and Evolutionary Computation*, GECCO '07, page 2945–2952, New York, NY, USA, 2007. Association for Computing Machinery.

[144] R. E. Smith and Max Kun Jiang. A learning classifier system with mutual-information-based fitness. In *2007 IEEE Congress on Evolutionary Computation*, pages 2173–2180, 2007.

[145] Lale Özbakir, Adil Baykasoğlu, and Sinem Kulluk. A soft computing-based approach for integrated training and rule extraction from artificial neural networks: Difaconn-miner. *Applied Soft Computing Journal*, 10:304–317, 1 2010.

[146] Ruxandra Stoean and Catalin Stoean. Modeling medical decision making by support vector machines, explaining by rules of evolutionary algorithms with feature selection. *Expert Systems with Applications*, 40:2677–2686, 6 2013.

[147] Ivanoe De Falco. Differential evolution for automatic rule extraction from medical databases. *Applied Soft Computing Journal*, 13:1265–1283, 2013.

[148] Jinkun Wang, Yezheng Liu, Jianshan Sun, Yuanchun Jiang, and Chunhua Sun. Diversified recommendation incorporating item content information based on moea/d. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 688–696, 2016.

[149] Amr Alzouhri Alyafi, Monalisa Pal, Stephane Ploix, Patrick Reignier, and Sanghamitra Bandyopadhyay. Differential explanations for energy

management in buildings. In *2017 Computing Conference*, pages 507–516, 2017.

[150] Amr Alzouhri Alyafi, Jean-Phillipe Guillbaud, Patrick Reignier, and Stephane Ploix. Explanations engine for energy management systems in buildings. In *2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, volume 2, pages 722–729, 2017.

[151] Zhiwen Xiao and Jianbin Jiao. Interpretable credit risk assessment based on heuristic knowledge extraction method. In *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 1237–1241, 2021.

[152] Youcong Ni, Xin Du, Peng Ye, Leandro L. Minku, Xin Yao, Mark Harman, and Ruliang Xiao. Multi-objective software performance optimisation at the architecture level using randomised search rules. *Information and Software Technology*, 135, 7 2021.

[153] Iztok Fister and Iztok Fister. Information cartography in association rule mining. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 6:660–676, 6 2022.

[154] João Roberto Bertini Junior and Alberto Cano. An explainable classifier based on genetically evolved graph structures. In *2022 IEEE Congress on Evolutionary Computation (CEC)*, pages 01–08, 2022.

[155] Md Meftahul Ferdaus, Ripon K. Chakrabortty, and Michael J. Ryan. Multiobjective automated type-2 parsimonious learning machine to forecast time-varying stock indices online. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52:2874–2887, 5 2022.

[156] Alberto Segura-Delgado, Augusto Anguita-Ruiz, Rafael Alcalá, and Jesús Alcalá-Fdez. Mining high average-utility sequential rules to identify high-utility gene expression sequences in longitudinal human studies. *Expert Systems with Applications*, 193, 5 2022.

[157] Chaima Abid, Dhia Elhaq Rzig, Thiago Do Nascimento Ferreira, Marouane Kessentini, and Tushar Sharma. X-sbr: On the use of the

history of refactorings for explainable search-based refactoring and intelligent change operators. *IEEE Transactions on Software Engineering*, 48:3753–3770, 10 2022.

[158] Ivanoe De Falco, Giuseppe De Pietro, and Giovanna Sannino. A two-step approach for classification in alzheimer's disease. *Sensors*, 22, 6 2022.

[159] David Pätzel, Michael Heider, and Jörg Hähner. Towards principled synthetic benchmarks for explainable rule set learning algorithms. In *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, GECCO '23 Companion, page 1657–1662, New York, NY, USA, 2023. Association for Computing Machinery.

[160] Angela Troncoso Garcia, Alicia Troncoso Lora, Francisco Martinez-Alvarez, and Maria Martinez Ballesteros. Evolutionary computation to explain deep learning models for time series forecasting. In *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing*, SAC '23, page 433–436, New York, NY, USA, 2023. Association for Computing Machinery.

[161] Ivanoe De Falco, Giuseppe De Pietro, and Giovanna Sannino. Classification of covid-19 chest x-ray images by means of an interpretable evolutionary rule-based approach. *Neural Computing and Applications*, 35:16061–16071, 8 2023.

[162] Giovanni Misitano. Exploring the explainable aspects and performance of a learnable evolutionary multiobjective optimization method. *ACM Transactions on Evolutionary Learning and Optimization*, 3 2023.

[163] Salvatore Corrente, Salvatore Greco, Benedetto Matarazzo, and Roman Słowiński. Explainable interactive evolutionary multiobjective optimization. *Omega*, 122, 1 2024.

[164] Muhammed Ulucan, Gungor Yildirim, Bilal Alatas, and Kursat Esat Alyamac. A new intelligent sunflower optimization based explainable artificial intelligence approach for early-age concrete compressive strength classification and mixture design of rac. *Structural Concrete*, 24:7400–7418, 12 2023.

[165] Majid Almaraash, Mahmoud Abdulrahim, and Hani Hagras. A lifelong learning xai metaheuristic-based type-2 fuzzy system for solar radiation modelling. *IEEE Transactions on Fuzzy Systems*, 2023.

[166] Ali Danandeh Mehr, Vahid Nourani, Ercan Kahya, Bahrudin Hrnjica, Ahmed M.A. Sattar, and Zaher Mundher Yaseen. Genetic programming in water resources engineering: A state-of-the-art review. *Journal of Hydrology*, 566:643–667, 11 2018.

[167] Leonardo Lucio Custode and Giovanni Iacca. Interpretable pipelines with evolutionary optimized modules for reinforcement learning tasks with visual inputs. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, GECCO '22, page 224–227, New York, NY, USA, 2022. Association for Computing Machinery.

[168] Marco Virgolin, Andrea De Lorenzo, Francesca Randone, Eric Medvet, and Mattias Wahde. Model learning with personalized interpretability estimation (ml-pie). In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, GECCO '21, page 1355–1364, New York, NY, USA, 2021. Association for Computing Machinery.

[169] Joe Harrison, Tanja Alderliesten, and Peter A. N. Bosman. Gene-pool optimal mixing in cartesian genetic programming. In Günter Rudolph, Anna V. Kononova, Hernán Aguirre, Pascal Kerschke, Gabriela Ochoa, and Tea Tušar, editors, *Parallel Problem Solving from Nature – PPSN XVII*, pages 19–32, Cham, 2022. Springer International Publishing.

[170] Binh Tran, Chamika Sudusinghe, Su Nguyen, and Damminda Alahakoon. Building interpretable predictive models with context-aware evolutionary learning. *Applied Soft Computing*, 132, 1 2023.

[171] Bogdan Burlacu, Kaifeng Yang, and Michael Affenzeller. Population diversity and inheritance in genetic programming for symbolic regression. *Natural Computing*, 2023.

[172] Dale I. Foreman Gregory W. Corder. *Testing Data for Normality*, chapter 2, pages 12–37. John Wiley & Sons, Ltd, 2009.

[173] Amitava Mitra. The taguchi method. *WIREs Computational Statistics*, 3(5):472–480, 2011.

[174] F. W. Scholz and M. A. Stephens. K-sample anderson-darling tests. *Journal of the American Statistical Association*, 82(399):918–924, 1987.

[175] William H. Kruskal and W. Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621, 1952.

[176] Giulia Vilone and Luca Longo. Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion*, 76:89–106, December 2021.

[177] Felix Lanfermann, Sebastian Schmitt, and Patricia Wollstadt. Understanding concept identification as consistent data clustering across multiple feature spaces. In *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 180–189, 2022.

[178] Felix Lanfermann, Qiqi Liu, Yaochu Jin, and Sebastian Schmitt. Identification of energy management configuration concepts from a set of pareto-optimal solutions. *Energy Conversion and Management: X*, 22:100576, 2024.

[179] Aidan Wallace, Alexander E. I. Brownlee, and David Cairns. Towards Explaining Metaheuristic Solution Quality by Data Mining Surrogate Fitness Models for Importance of Variables. In Max Bramer and Richard Ellis, editors, *Artificial Intelligence XXXVIII*, volume 13101, pages 58–72. Springer International Publishing, Cham, 2021. Series Title: Lecture Notes in Computer Science.

[180] Fernando Lezama, Joao Soares, Ricardo Faia, Zita Vale, Olli Kilkki, Sirpa Repo, and Jan Segerstam. Bidding in local electricity markets with cascading wholesale market integration. *International Journal of Electrical Power & Energy Systems*, 131:107045, 2021.

[181] Yoan Martínez-López, Ansel Y. Rodríguez-González, Julio Madera, Miguel Bethencourt Mayedo, and Fernando Lezama. Cellular estimation of distribution algorithm designed to solve the energy resource management problem under uncertainty. *Engineering Applications of Artificial Intelligence*, 101:104231, May 2021.

[182] Shintaro Ikeda and Ryozo Ooka. Metaheuristic optimization methods for a comprehensive operating schedule of battery, thermal energy

storage, and heat source in a building energy system. *Applied energy*, 151:192–205, 2015.

[183] Fernando Lezama, Joao Soares, Ricardo Faia, Tiago Pinto, and Zita Vale. A New Hybrid-Adaptive Differential Evolution for a Smart Grid Application Under Uncertainty. In *2018 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–8, Rio de Janeiro, July 2018. IEEE.

[184] Mohamed A Mohamed, Ali M Eltamaly, and Abdulrahman I Alolah. Swarm intelligence-based optimization of grid-dependent hybrid renewable energy systems. *Renewable and Sustainable Energy Reviews*, 77:515–524, 2017.

[185] João Soares, Fernando Lezama, Ricardo Faia, Steffen Limmer, Manuel Dietrich, Tobias Rodemann, Sergio Ramos, and Zita Vale. Review on fairness in local energy systems. *Applied Energy*, 374:123933, November 2024.